DOCUMENT RESUME

ED 319 185 EC 230 874

AUTHOR Treece, Nelson L.
TITLE Use of the Cloze Procedure To Measure Reading Comprehension and Language Ability of the Deaf.
PUB DATE Jun 89
NOTE 89p.
PUB TYPE Information Analyses (070)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Cloze Procedure; Curriculum Design; *Deafness; Elementary Secondary Education; Evaluation Methods; *Reading Comprehension; Reading Skills; *Reading Tests; Second Language Learning; Student Evaluation; Teaching Methods; *Test Validity

ABSTRACT

This literature review examines use of the cloze procedure for assessing reading and language competence and for planning individualized reading curricula. The first half of the paper describes studies performed with non-hearing-impaired students, discussing validation of the cloze procedure, establishment of criterion test scores, and the influence of various deletion patterns and manipulation of larger contexts. The second half covers performance by deaf subjects using the cloze procedure. Findings indicate that while cloze methods are valid for use with both native-speaking and second-language populations, its validity for reading assessment remains uncertain, and the technique cannot yet be recommended for teachers of the deaf seeking to match readers with text. Includes over 200 references. (PB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED319185

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to:

In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

USE OF THE CLOZE PROCEDURE TO MEASURE READING COMPREHENSION AND LANGUAGE ABILITY OF THE DEAF

QUALIFYING PAPER
SUBMITTED BY

NELSON L. TREECE

JUNE, 1989

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Nelson L Treece

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EC 230874

BEST COPY AVAILABLE

TABLE OF CONTENTS

1.0 Introduction.....1

2.0 Overview.....3

STUDIES WITH NON-HEARING-IMPAIRED SUBJECTS

3.0 Description of the Cloze Procedure.....5
- 3.1 Deletion System.....5
- 3.2 Scoring System.....7
- 3.3 Passage Length.....8

4.0 Practical Validation of the Cloze Procedure.....9
- 4.1 Ranking of Text.....10
- 4.2 Correlation with Traditional Tests.....11
- 4.3 Establishing Criterion Cloze Scores.....13
- 4.4 Summary.....16

5.0 What Does the Cloze Procedure Measure?.....17
- 5.1 Test Format.....17
- 5.2 Factor Analyses.....21
- 5.3 Manipulation of Deletion Patterns.....24
- 5.4 Manipulation of Larger Contexts.....25
- 5.5 Summary.....27

6.0 A Perspective on the Cloze Procedure.....28
- 6.1 The Dual Function of Cloze.....28
- 6.2 A Perspective on the Literature.....34
- 6.3 Summary.....36

STUDIES WITH HEARING-IMPAIRED SUBJECTS

7.0 Deaf Performance on the Cloze Procedure.....38
- 7.1 Test Format.....39
- 7.2 Manipulation of Deletion Patterns.....40
- 7.3 Manipulation of Larger Contexts.....41
- 7.4 Ranking of Text.....42
- 7.5 Correlation with Traditional Tests.....46
- 7.6 Establishing Criterion Cloze Scores.....50
- 7.7 Summary.....52

8.0 Summary of Findings.....52

9.0 Implications for Deaf Education.....54

10.0 Directions for Future Research.....57

REFERENCE NOTES.....58

REFERENCES.....65

APPENDIX.....79

Table 1: Ranking Studies.....80

Table 2: Correlations of Cloze Performance by Native Speakers with Same-Passage Multiple-Choice Measures.....81

Table 3: Correlations of Cloze Performance by Native Speakers with Standardized Reading Measures.....82

Table 4: Correlations of Cloze Performance by Second Language Learners with Reading Comprehension Measures.....83

Table 5: Correlations of Cloze Performance by Second Language Learners with Language Proficiency Measures.....84

Table 6: Criterion Cloze Score Studies.....85

1.0 Introduction

The cloze procedure, in which subjects are asked to supply words that have been systematically deleted from running text, was first introduced to educators by Taylor (1953) as a way of measuring text difficulty.[1] Since that time, extensive research has shown that the cloze procedure has great utility for measuring various aspects of reading and language competence as well. Cloze tests have repeatedly been demonstrated to be valid and reliable instruments with which to measure the reading comprehension ability of native speakers and the foreign language proficiency of second language learners (Buros 1978). Well-replicated findings, coupled with the ease and objectivity of cloze test construction (Bormuth 1975) and the unique capacity of these tasks to directly access the interaction of reader with text (Spolsky 1969), recommend cloze tests highly for use as an assessment device with a wide range of subjects.

One of the most prevalent uses of the cloze procedure in the field of reading instruction has been as a practical instrument by which classroom teachers may match students with text appropriate to their abilities (Bormuth 1975). The extensive calculations often required by standard readability formulae (for example, see Dale & Chall 1948), and the investment of time involved in the administration of an informal reading inventory (for example, see Betts 1946), can be avoided by the classroom teacher if the cloze procedure is used in their stead to accomplish such a match. It is this practical, instruction-focussed application of

the cloze procedure which is often featured in texts written for teachers of reading (for example, see Harris & Sipay 1985; NY City Board of Education 1982).

Teachers of the deaf also make use of the cloze procedure to match readers with text (LaSasso 1978a). The limited body of research which has examined the cloze procedure in relation to the deaf population, however, suggests that caution should be exercised in this regard. Cloze scores of deaf students appear not to reflect levels of text difficulty (LaSasso 1978b, 1980) in the same way that cloze scores of hearing students do (Bormuth 1963a; Taylor 1953). In addition, hearing and deaf students with otherwise comparable reading comprehension ability perform very differently on identical cloze passages (LaSasso 1978b, 1980; Marshall 1970a,b; Moores 1967, 1970). Discrepancies between the performance of deaf and hearing subjects on cloze tasks have in fact led LaSasso (1978b, 1980) to question the validity of the cloze procedure as a measure of reading comprehension for the deaf.

The problem for teachers of the deaf is clear: given the success of the cloze procedure in measuring reading competence in other populations and the apparently aberrant findings for the deaf, to what extent can these teachers properly make use of this technique as a practical device for matching students with text in their classrooms? This qualifying paper will examine the existing research literature in an attempt to answer this question.

2.0 Overview

Two distinct but parallel traditions of research with non-hearing-impaired populations inform our understanding of the cloze procedure and its potential for use with the deaf - those studies investigating the cloze procedure as a measure of reading comprehension, and those examining cloze as a language assessment device. In the first fifteen years of cloze research, beginning with Taylor's 1953 publication, studies focussed exclusively on cloze as a measure of readability and reading comprehension, although it was suggested as early as 1956 by Taylor that cloze could have application as a language proficiency measure as well. Aside from a study by Carroll *et al.* (1959), no investigations of the cloze procedure as a measure of linguistic ability appeared in the literature until Darnell (1968) and Spolsky (1969) a decade later.

The reading research and the language proficiency traditions have largely focused on separate populations. Studies concerned with the cloze procedure as a measure of language ability have worked exclusively with hearing persons learning a second language. Researchers looking at the cloze procedure as a measure of reading comprehension have concentrated overwhelmingly on hearing native speakers, although evidence has also come from the performance of second language learners. While these two types of inquiry are distinct, an understanding of both is necessary in order to draw as complete and accurate a picture as possible of the cloze procedure.

In order to set the stage for a consideration of deaf performance on cloze tasks, this qualifying paper will examine the findings of these two research traditions. A brief description of the cloze procedure will be given in section three, including a discussion of deletion patterns, scoring systems, passage length, and reliability. Next, the three types of research findings which have been used to validate the cloze procedure as a practical measure of reading comprehension and language proficiency will be revieved in section four. In section five ve will take a closer look at the issue of what the cloze procedure is actually measuring, and explore the disagreements and discrepancies which have arisen in this body of research. An interpretive framevork will be proposed in section six, within which these apparent contradictions may be resolved. Data about the performance of deaf subjects on cloze tasks will then be revieved in section seven, and compared to the performance of non-hearing-impaired populations. Specific suggestions for research design will be proposed for each type of study. Section eight will present a summary of findings, and section nine will discuss the implications of this reviev of literature for the use of the cloze procedure with the deaf. Lastly, directions which future research should take in order to clarify the role which the cloze procedure can play in the assessment and instruction of the deaf population will be offered in the tenth and final section.

3.0 Description of the Cloze Procedure

The standard cloze format requires subjects to replace words missing in text without the accompaniment of prompts or distractors.[2] There are three primary dimensions along which such tests can vary. The first dimension is the deletion scheme used to remove words from the passage. The second is the scoring system employed to compute the percentage of correct responses. The final variable, the most important factor in determining the reliability of this procedure, is the amount of text used in constructing the cloze passage (correspondingly, the number of items which make up the test).

3.1 Deletion System

The "standard procedure" for constructing cloze tests for the purpose of reading research with native speakers has long been the pattern of replacing every 5th word in a text, regardless of word type, with a blank of uniform length (Taylor 1953; Bormuth 1975). The research investigating second language competence has tended to resort to an every 7th word deletion scheme (e.g., Briere & Hinofotis 1979; Caulfield & Smith 1981; Hanzeli 1977; Hinofotis 1980; Irvine *et al.* 1974; Oller 1972; Oller & Conrad 1971). These numbers were arrived at only after researchers had experimented with a variety of deletion patterns.

The most influential and widely accepted rationale for the spacing of deletions was published by MacGinitie (1961). In a study focussing on the effects of deleting every 3rd, every 6th, every 12th, or every 24th word, he

found that cloze scores did not differ significantly when the amount of context surrounding the deleted word was five words or more, but that scores dropped significantly when only two words were left in between deletions. As a result of these and other similar findings for both native and second language speakers (Aborn et al. 1959; Alderson 1979c; Fillenbaum et al. 1963; Haskell 1973 a,b; McNinch et al. 1974; Miller & French 1974; Porter 1983; Salzinger et al. 1962; Taylor 1953), deletion patterns ranging from every 5th on up to every 12th word are found scattered throughout the literature, and patterns involving every 3rd or 4th word are avoided.

In addition to allowing any type of word to be subject to deletion, early cloze researchers (for example, see Greene 1965a,b; Rankin 1958, 1959a; Taylor 1956, 1957; Weaver & Kingston 1963) experimented with an alternate way of constructing cloze tests which specified the nature of the words to be dropped, e.g., only content words. Taylor (1956, 1957) concluded from his studies that the "any-word" system was superior for two reasons: for most correlational purposes it provided an equal or superior measure, and it gave a better proportional representation of the word types making up the original content of the writer's message. Common agreement has established use of the "any-word" deletion scheme as the standard in constructing cloze tests for both native and second language populations (Bormuth 1975; Oller 1979).

3.2 Scoring System

Early investigators of the cloze procedure experimented with various scoring systems (for example, see Bormuth 1963b, 1965b; Fillenbaum & Jones 1962; Gallant 1965; Hafner 1964; Miller & Coleman 1967; Rankin 1958, 1959a; Ruddell 1964a,b; Taylor 1953). The most common types were verbatim scoring, that is, counting only exact word replacements as correct, and "acceptable alternate" scoring, in which responses that make sense given the grammar and context, but don't necessarily match the author's words exactly, are counted as correct. These studies found that both methods resulted in virtually identical correlations and text rankings for native speaking readers (one study, Miller & Coleman 1967, found that the two scoring systems correlated at .99). Because verbatim scoring is able to avoid problems of subjective scorer judgment and is much more efficient to employ, this scoring method became the standard procedure for measuring the reading comprehension of native speakers (Bormuth 1975).

When researchers began to investigate the cloze procedure as a measure of the foreign language proficiency of second language learners, however, they found themselves in disagreement about which scoring system to use. Some research found no advantage to the extra effort required to use acceptable alternate scoring (e.g., Anderson 1972; Haskell 1973a,b; Irvine et al. 1974; Stubbs & Tucker 1974). Other studies found that verbatim scoring was too severe a measure for second language learners, and that the increased

variance in scores which resulted from employing acceptable alternate scoring led to healthier correlations and higher reliability (e.g., Alderson 1979a,b; Brown 1980; Hinofotis 1980; Oller 1972; and see Appendix, Table 5). As a result, both types of scoring are well represented in the second language literature (Brown 1983).[3]

3.3 Passage Length

Bormuth (1964, 1965a, 1975) established that the length of text necessary for adequate reliability with the cloze procedure is a format which includes at least 50 items. For an every 5th word deletion scheme, for example, a selection of at least 250 words would be required (in addition to the first and last sentences of the passage, which are commonly left intact). Including 50 items on the test yields a reliability coefficient of around .85, which Bormuth (1975) concludes should be adequate for most purposes. Including substantially fewer items drastically reduces the reliability of the results (Bormuth 1964), while it would take a passage of at least 100 items (a minimum of 500 words using every-5th deletion) to improve reliability but little, to about .90 (Bormuth 1975). In order to maximize reliability when matching reader with text, Bormuth (1964) urges that either more items be used (by selecting several passages from the text, for example) or that all possible cloze forms over a single passage be constructed and administered (i.e. using an every-5th system, the first cloze form would delete words 1, 6, 11, etc., the second form words 2, 7, 12, etc., and so on).

Practical validation studies adhering to these length specifications have consistently reported reliability coefficients for native speakers (Potter 1968), and for second language subjects across both popular scoring methods (Brown 1980), that are comparable to Bormuth's figures.

4.0 Practical Validation of the Cloze Procedure

Three types of research have been used with non-hearing-impaired populations to validate the cloze procedure as a practical measure of reading comprehension and language proficiency: 1) comparison with independent text rankings; 2) correlation with standard tests; and 3) the establishment of criterion levels of cloze performance.

The first type of research, the ranking studies, has been conducted solely to explore cloze as a measure of reading comprehension, with four second language groups contributing data (Alderson 1980; Anderson 1972; Haskell 1973a,b; Oller 1972). The second type of research, correlation with standard tests, has been extensively applied to cloze investigations both of reading comprehension and of language proficiency. Second language learners, in addition to providing all the data for the language studies, are well-represented in the reading correlations as well. The establishment of criterion cloze scores, the third type of research to be examined here, has been widely undertaken only in relation to cloze as a measure of reading comprehension - specifically, with the goal of attaching meaningful interpretations to cloze scores for the purpose of matching readers with text in the

classroom. Only one study has dealt with this use of cloze with second language learners (Anderson 1971).

4.1 Ranking of Text

Two techniques have been employed to investigate the ability of cloze tasks to rank passages in the same manner as an independent measure (see Appendix, Table 1). These studies either compare the ranking of mean cloze scores with the ranking of the same passages by readability formulae or expert judgment of text difficulty (Alderson 1980; Anderson 1972; Bormuth 1963a; Entin & Klare 1978; Froese 1971; Haskell 1973a,b; Nestvold 1972; Oller 1972; Robinson 1972; Taylor 1953), or compare the ranking of mean cloze scores with the scores the same subjects receive on multiple-choice tests constructed from the same materials (Bormuth 1963a, 1968; Entin & Klare 1978).

In only three cases do these studies fail to demonstrate comparable text ranking by cloze scores and independent criteria, and in each instance there are mitigating conditions which may account for these results. Froese (1971) found a significant but small negative correlation (r=-.22) between cloze scores on randomly selected passages from sixth grade science textbooks and the ranking of these passages by the Dale-Chall readability formula. Froese suggests, however, that the specialized scientific vocabulary used in these passages may have thrown off the accuracy of the readability formula, resulting in an inappropriate ranking of the texts. The cloze scores of Robinson's (1972) subjects, functionally illiterate prison

inmates, did not rank passages in the same way as did several readability formulae, but he points out that previous research shows readability formulae not to be successful at rating text for illiterate adult populations. Entin and Klare (1978) found no significant correlation between cloze score ranking and their criterion (multiple-choice scores on passages from the Nelson-Denny Reading Test), but they suggest that peculiarities inherent in the Nelson-Denny format may have accounted for this.

Conclusion. Overall, the ranking studies suggest that the cloze test is able to measure reading comprehension, at least in the sense that mean cloze scores reflect the difficulty level of text as determined by readability formulae or expert judgment. This has been found both for native and for second language subjects.

4.2 Correlation with Traditional Tests

Reading Comprehension. At least fifteen studies have sought correlations between cloze scores and independent measures of reading comprehension for native speakers. These studies either figure correlations between cloze scores and scores on multiple-choice tests constructed over the same passages (Bormuth 1963a, 1967; Entin & Klare 1978; Jenkinson 1957; Rankin 1959a; Rankin & Culhane 1969; Taylor 1956), or compute correlations between cloze scores and scores on standardized measures of reading comprehension (Fletcher 1960; Froese 1971; Gallant 1965; Greene 1965a,b; Jenkinson 1957; Rankin 1959a; Ruddell 1964a,b; Schneyer 1965; Schoelles 1971; Weaver & Kingston 1963). In addition, at

least ten studies have examined correlations between second language cloze performance and either standardized reading tests (Flahive 1980; Friedman 1964) or reading subtests included in foreign language measures (Alderson 1979a,b; Bachman 1985; Hinofotis 1980; Irvine et al. 1974; Oller 1972; Oller & Conrad 1971; Pike 1979; Stubbs & Tucker 1974).

For native speakers, correlations between cloze scores and multiple-choice scores over the identical, intact passages (see Table 2) range in strength from r=.57 to r=.95, with most values of rho falling in the .70's and .80's. No study failed to find a significant correlation with the criterion. When cloze performance is compared with scores on standardized measures of reading comprehension (see Table 3), values of rho range from .25 to .85, with most reported values falling in the .60's and .70's. Correlations for second language learners (see Table 4) range from .47 to .87, with most values in the .60's and .70's.

Language Proficiency. Several widely accepted measures of second language proficiency (for example, the Test of English as a Foreign Language [TOEFL]) have been used as independent criteria against which to measure cloze performance. At least fourteen studies have sought correlations in this manner (Alderson 1979a,b; Bachman 1985; Briere & Hinofotis 1979; Carroll et al. 1959; Caulfield & Smith 1981; Flahive 1980; Hanania & Shikhani 1986; Heilenman 1983; Hinofotis 1980; Irvine et al. 1974; Mullen 1979b; Oller 1972; Oller & Conrad 1971; Stubbs & Tucker 1974). As

can be seen from an examination of Table 5, correlations range from r=.55 to r=.88 using verbatim scoring, and from r=.62 to r=.90 using acceptable alternate scoring. Most values of rho fall in the .70's and .80's.

Conclusion. The overall magnitude of these correlations is strong evidence that the cloze procedure is able to measure both reading comprehension and language ability, at least in the sense that cloze tests and standard assessment devices appear to largely be measuring the same thing.

If one examines the correlations in Tables 2 and 3, it becomes clear that cloze scores correlate more highly with multiple-choice tests constructed over the same passage than they do with standardized reading measures for native speakers (cf. Rankin 1965). This would suggest that the cloze procedure may be suitable as a substitute for multiple-choice testing in certain classroom situations. The attempt to validate one such use of cloze will be examined in the following section.

4.3 Establishing Criterion Cloze Scores

Reading Comprehension. This final body of research holds the key to the practical application of the cloze procedure in the classroom. These investigations sought to go beyond the mere fact of correlation and establish levels of cloze performance which the reading teacher can use to match students with appropriate text.

This research took as its reference points the multiple-choice criterion scores proposed by Betts (1946) for determining the independent (>90% correct),

instructional (75% - 90% correct), and frustration (<50% correct) reading levels of texts. The instructional range is of particular interest to classroom teachers. Texts on which students are able to score between 75% and 90% correct when given a multiple-choice test over the same passage are said to pose just the degree of difficulty and challenge necessary for students to benefit from them with guidance from the classroom teacher (Betts 1946, pp. 447-450). Researchers have thus tried to isolate cloze scores that could consistently predict 75% - 90% correct responses on same-passage multiple-choice tests.

At least twelve studies have sought explicitly to establish criterion cloze scores[4] (see Table 6). These studies may be grouped into four broad categories: those employing regression procedures to determine equivalences (Alexander 1969; Bormuth 1967; Entin & Klare 1978; Jones & Pikulski 1974; Rankin & Culhane 1969); those using matching techniques (Almeida 1975; Bormuth 1968); those figuring cloze scores for grade-appropriate material (Peterson & Carroll 1974; Peterson *et al.* 1972, 1973); and one study which observes the amount of information gain over the entire range of cloze scores (Bormuth 1969a). The sole researcher to establish criterion cloze scores for second language learners (Anderson 1971) does not report the statistical method used in his study.

Despite the variety of techniques employed, types of material, ages of subjects, and language experience, surprisingly consistent levels of cloze scores emerge which

correspond to the instructional range. For the 75% level on multiple-choice comprehension tests, cloze scores ranging from 31% to 47% have been found, with most values falling in the low to mid 40's. For the 90% level on multiple-choice comprehension tests, cloze scores ranging from 36% to 61% were found, with most values falling in the mid to upper 50's. The one study which found highly divergent cloze equivalents (Entin & Klare 1978, instructional range from 31% to 36%) allowed students, in contrast to other criterion studies, to refer back to the passage while completing the multiple-choice items. They suggest that this artificially inflated the multiple-choice scores compared to other studies, resulting in lower criterion cloze scores than would otherwise have been obtained. Bormuth's (1968) cloze equivalents for the instructional range (44% - 57%) are the scores commonly reported as representative of this research (for example, see Harris & Sipay 1985; King & Quigley 1985; NY City Board of Education 1982).

Language Proficiency. No attempts could be found in the literature to establish criterion cloze scores for second language students in relation to the various tests of foreign language proficiency available.[8] Studies using the TOEFL as their criterion, for example, did not seek to isolate cloze scores that consistently correspond to established levels of achievement on that test. The applicability of such criterion cloze scores would be limited, in any case, since they would be useful only to those programs making use of the specific proficiency tests

in question. In contrast, the criterion cloze scores established for reading comprehension, based as they are on multiple-choice tests constructed over the very same intact passages, have potential applicabillity in any classroom using written text for instruction.

Conclusion. These criterion studies have been successful in establishing the basis for using the cloze procedure to match readers with appropriate levels of text in the classroom, in the sense that cloze scores have been found which would seem to carry the same interpretation for the reading teacher as do well-established levels of multiple-choice performance. The findings apply primarily to native language speakers, although one study has confirmed these criterion scores for second language students as well.

4.4 Summary

The evidence seems overwhelming in support of the cloze procedure as a measure of both reading comprehension and language ability. Yet, correlational studies alone leave many questions unansvered concerning the nature of the cloze procedure. While these studies can tell us a great deal about how the cloze procedure appears to be similar to other assessment devices, they are much less capable of specifying those features which are unique to the cloze procedure itself. That is, since cloze scores do not correlate perfectly with independent criteria, some variance attributable to the cloze procedure always remains unaccounted for and needs to be explained. A more precise characterization of what the cloze procedure measures would

thus be desirable, and so we turn next to those studies which have spoken directly to this issue.

5.0 What Does the Cloze Procedure Measure?

The results of these inquiries, in contrast to the practical validation studies revieved in section 4.0, have given rise to much disagreement and difference of opinion in both the reading comprehension and the language proficiency camps.

This debate has been fueled by four sources. Research has addressed the question first of all through theoretical arguments concerning the item format of the cloze test. Factor analytic techniques have also been employed in an attempt to specify what the cloze procedure actually measures. Another group of studies, which examines the effect of various deletion patterns on cloze performance, has contributed to the controversy as well. Finally, investigations manipulating the amount and quality of context in and around cloze passages have yielded data relevant to the issue of what the cloze procedure is measuring.

We will examine each type of evidence in turn. After completing the reviev of these studies, I will suggest an interpretive framevork in section 6.0 to help explain why such apparent contradictions should arise. The perspective gained from this account will also prepare the way for a consideration of the cloze procedure and the deaf.

5.1 Test Format

Reading Comprehension. A feature of the cloze test that

is highlighted by proponents as one of its greatest strengths, namely, the ease and objectivity of item selection, has also engendered some of the greatest criticism of this test format.

Cloze test design does have some outstanding virtues. Cloze tests can be quickly constructed by teachers in a clerical fashion, with no special expertise in test design or interpretation required, and without the subjectivity and variability which inevitably accompanies the writing of comprehension questions (Bormuth 1975). Because cloze items are selected randomly and not subject to the judgment of test constructors, they offer an unbiased set of items which are representative of all facets of the passage targeted for comprehension (Rankin 1978; Simons 1971; Taylor 1953). The cloze test, in these regards, clearly avoids some of the problems inherent in other reading comprehension measures.

But researchers also report (for example, see Rankin 1978) that this very design advantage raises serious concerns for some regarding the test format of the cloze procedure. Namely, the items on a cloze test do not appear to be demanding the same sort of responses which have traditionally been thought of as reflecting reading comprehension, e.g., they don't ask the reader to choose a main idea, recount facts, interpret a sequence of events, perceive the author's purpose, paraphrase content, etc. Common sense and expert opinion alike might ask how reading comprehension could be assessed by the cloze procedure if the test items fail to elicit what has traditionally been

considered indicative of the reading comprehension process (cf. Rankin & Helm 1986).

This criticism seems telling, until one realizes that the appropriateness of items on traditional comprehension tests are themselves open to question (e.g., see Carroll 1972; Carver 1970; Simons 1970, 1971; Tuinman 1972). In his critique of attempts to measure reading comprehension, Simons (1971) points out that traditional tests measure, not the process of comprehension itself, but rather behaviors from which we can only infer that comprehension has taken place, with all sorts of extraneous factors thrown in to interfere with the precision of the measurement. Reading research lacks a coherent theory of what reading comprehension itself is, he argues, and thus it is not possible to specify just what items would most appropriately measure the underlying comprehension process. He opts for use of the cloze procedure over traditional comprehension tests because, even though it is open to criticism, cloze at least seems to allow less interference from extraneous factors than do more traditional reading comprehension measures (Simons 1971).

It is this directness of cloze tasks, the way cloze items are able to assess comprehension during the actual reading of text, that provides reading researchers with their strongest rationale for the cloze test format (for example, see Rankin 1978; Taylor 1953). For Taylor, "Cloze procedure takes a measurement of the likeness between the patterns a writer has used and the patterns the reader is

anticipating while he is reading (1953, p. 417)." For Rankin, "... all cloze response[s] are based upon the basic psycholinguistic process of inference which is intrinsic to all communication ... It is precisely this process that is tapped by all cloze items in varying degrees. This cannot be said of conventional comprehension measurements (1978, p. 151)." Accepting the premise that reading tests should measure the comprehension process as directly as possible, these researchers present the cloze procedure as eminently well-suited to this task.

Language Proficiency. The issue of cloze format has been less controversial in relation to measuring language proficiency than it has in measuring reading comprehension because, for one thing, cloze test items are similar to many items commonly included in tests of foreign language competence (for example, see the Structure and Written Expression section of the TOEFL Sample Test, ETS 1985). Second language researchers also argue forcefully that those skills which are required to successfully complete cloze items, that is, the ability to formulate correct hypotheses about language based on familiarity with its syntactic patterns and semantic relationships, are also those skills that best reveal the degree to which second language learners have mastered the target native language (Aitken 1977; Oller 1973; Spolsky 1969).

These researchers also point to evidence that cloze items are able to tap not only structural knowledge, but highly integrative language abilities as well, in the same

manner as do listening comprehension, reading comprehension, or dictation tests (Irvine *et al.* 1974; Oller & Conrad 1971). They even invoke, in the spirit of their reading counterparts, the psycholinguistic model of reading put forvard by Goodman (1971) to describe the active hypothesis testing and confirming process to which, they feel, the cloze procedure is privy (Aitken 1977; Oller 1973).

Conclusion. Researchers seem largely to agree that the cloze procedure should be a viable way by which to measure both reading comprehension and language proficiency. Their answer to the theoretical question - what does the cloze procedure measure? - is that it taps the processes of comprehension and the skilled inferencing with language which are the proper objects of test assessment.

The empirical question as to what the cloze test is actually able to measure, even granted this theoretical acceptability, however, is a different issue entirely. Accordingly, the remainder of this section will be devoted to examining that research which has looked to performance data for evidence.

5.2 Factor Analyses

An extension of correlational methods, factor analysis attempts to identify the unities underlying correlations and reveal thereby the deeper regularities which characterize such relationships. Research findings from the factor analyses, all but one of which (Bachman 1982) used native language speakers, split down the middle in regard to what they find the cloze procedure to be measuring. One group of

studies concludes that the cloze procedure is primarily measuring what is characterized as overall verbal, or global comprehension, ability (Bachman 1982; Bormuth 1969b; Horton 1973, 1974-75; Kohler 1966; Ohnmacht et al. 1970).

A second group of studies, however, arrives at quite a different interpretation (Carver et al. 1971-72; Ohnmacht & Fleming 1971; Weaver & Kingston 1963). Weaver and Kingston (1963) found that, unlike comprehension tests, cloze tests loaded only moderately on a verbal comprehension factor; rather, "the underlying ability here seems to be the recognition of redundancy characteristics of language. This factor is called therefore 'redundancy utilization' (p. 259)." Ohnmacht and Fleming (1971) reported that cloze tests loaded heavily on what they call a general cloze factor, rather than on a verbal ability factor. Carver et al. (1971-72) offered further support for the distinction between cloze and comprehension measures with their finding that cloze tests loaded mainly on a factor apparently related to the ability to restore deleted parts of a passage, which they called a cloze factor, rather than on a comprehension factor. They conclude that "...the cloze test used as a measure of comprehension includes a large component of variance unrelated to comprehension and quite specific to the cloze technique itself (1971-72, p. 49)."

That analyses should reveal a unique cloze or "redundancy utilization" factor is unsurprising. The cloze procedure clearly requires some skills distinct from traditional comprehension tests, skills related to the

ability to replace missing words in a text according to the redundancy of the language itself rather than (or in addition to) global comprehension of the message. What is damaging to the model of cloze test qua measure of reading comprehension is the suggestion that cloze is assessing this factor primarily of even exclusively, for one expects reading comprehension to encompass more than mere knowledge of structural redundancies. Just this sort of evidence led J.B. Carroll (1972), when revieving the findings on cloze, to conclude that the cloze procedure is

> ...dependent chiefly on what might be called the 'local redundancy' of a passage, i.e., the extent to which linguistic cues in the immediate environment (generally, in the same sentence) of a missing word tend to supply it...There is no clear evidence that cloze scores can measure the ability to comprehend or learn the major ideas or concepts that run through a discourse (pp. 18,19).

Findings of this sort are equally damaging to the model of cloze test qua measure of language proficiency, for proponents of the cloze procedure point not only to its ability to capture mastery of language redundancies as among its major strengths, but to its capacity to measure global, integrative language competence as well (e.g., Irvine et al. 1974). Research which shows that the cloze procedure does not tap language skills that stretch beyond the clause or sentence level severely weakens the claims of such cloze advocates.

Conclusion. The factor analyses yield conflicting results. No decision on these issues is possible, moreover, based on factor analysis alone. For one thing, while factor

analyses are able to provide powerful insights into the underlying structure of relationships, their results can also be highly variable depending on the statistical methods employed. In addition, since factor analyses are based on correlational patterns, converging evidence from experimental studies is required before the findings of these techniques can be confirmed as accurate. Studies manipulating the amount and quality of context surrounding cloze blanks represent a potential source of such evidence, and so their contribution to deciding this issue will nov be considered.

5.3 Manipulation of Deletion Patterns

In the first published study on cloze measurement, Taylor (1953) experimented with an every 5th vord, every 10th vord, and random 10% deletion patterns. He found that all three deletion patterns yielded equivalent scores. MacGinitie (1961) reported that deletion patterns permitting context beyond five vords did not result in better performance. With the exception of only one study (Nicol & Miller 1959), all other investigations (see section 3.1) either support these findings, or establish that deletion patterns somevhere between every 5th and every 10th vord provide the maximum context effect.

The conclusion one must drav from these findings, that increasing the context around a deleted vord does not improve one's cloze score after a certain point, runs counter to vhat one vould expect from a test of either reading or language competence. We vould hope, after all,

that such tests should be sensitive to relationships beyond the range of four or five words. The results of these studies, then, that subjects do not experience greater success in replacing missing words as stretches of text longer than a bare minimum are left intact between deletions, would seem to support those factor analyses which point to the importance of very local, intra-sentence context in the successful prediction of cloze items.

5.4 Manipulation of Larger Contexts

The influence of context beyond the sentence level on cloze scores has been examined by researchers using a number of different methods. While some come down on the side of intersentential influence, most conclude just the opposite.

Miller and Coleman (1967) examined whether there was a cumulative improvement in cloze scores as native speakers progressed farther into the passage. That is, they sought evidence for the influence of prior context on the ability to predict words. They found that there was "...apparently no increase in correct guesses from word 20 to word 140, and this indicates that guesses are constrained very slightly, if at all, by words from other sentences (p. 853)." Carroll et al. (1959) obtained similar results.

Some researchers have used the technique of scrambling the order of sentences in a cloze passage to determine the effect on native and second language performance. One group argues that the cloze test is sensitive to intersentential constraints, for reorganizing sentence order resulted in lower cloze performance when compared to the same passages

presented in coherent sequence (Carroll et al. 1959; Chihara et al. 1977; Cziko 1978; Darnell 1963; Oller 1975; Ramanauskas 1972; Rye 1984).

On the other hand, a second group of researchers using comparable scrambling techniques (Kibby 1980; Leys et al. 1983; Markham 1988; Shanahan & Kamil 1982, 1983; Shanahan et al. 1982) avers just the opposite, for they found no reduction in cloze scores under the scrambled condition. Shanahan and Kamil (1982, 1983) found especially strong evidence for the insensitivity of the cloze procedure to intersentential constraints. They report that when students were asked to read two versions of a cloze passage, one with intact and the other with scrambled sentence order, recall was much worse for the scrambled version. Recall was, in other words, clearly sensitive to intersentential constraints. When cloze scores on these scrambled and intact passages were compared, however, there was no statistical difference at all. This type of result led the authors to conclude that "...it seems to be unreasonable to use and interpret cloze in classroom practice as a global measure of reading comprehension (Shanahan et al. 1982, p. 250)."

A final type of investigation found that adding surrounding context to cloze passages led to no improvement in native speaker performance. These studies compared performance on cloze passages that were embedded in original, intact, text to scores on the same cloze passages presented in isolation (Erickson & Hansen 1974; Kibby 1980; Shanahan et al. 1982) or examined the effects of cumulative

context on cloze performance by allowing subjects to read longer and longer amounts of intact text before beginning to work on cloze passages (Bartoo 1975; Smith-Burke *et al.* 1978; Suhorsky 1975).

Conclusion. With the exception of half the scrambling studies, this body of research speaks with virtually one voice - context beyond the level of the individual sentence does not seem to influence performance on cloze tasks.

5.5 Summary

In contrast to the virtual unanimity of positive results obtained by the criterion-related research revieved in section 4.0, the disagreement among the studies considered in this section is quite striking. As soon as one moves beyond considering hov cloze tests correlate with (behave in a manner similar to) other standard measures in global terms, and attempts to specify more precisely the features which are unique to the cloze procedure itself, agreement between findings ceases to be common. On the one hand, certain types of evidence point to the cloze procedure as measuring overall verbal, or general comprehension, ability (half the factor analyses, section 5.2) and as being sensitive to intersentential or discourse level constraints (half the sentence scrambling studies, section 5.4). On the other hand, success on the cloze test is found by other studies to depend primarily on the ability to handle semantic and syntactic redundancies (half the factor analyses, section 5.2), redundancies which often are found not to extend beyond the clause or sentence level (deletion

studies, section 5.3; half the scrambling studies, studies in adding context, section 5.4).

Before considering the use of the cloze procedure with the deaf, we need somehow to come to terms with these discrepancies. What is needed is a framework for understanding the cloze procedure which can account for how such conflicting findings could arise, and suggest how they might ultimately be resolved. The following section of this qualifying paper will attempt to supply such a framework.

6.0 A Perspective on the Cloze Procedure

6.1 The Dual Function of Cloze

We will go a long way towards understanding the nature of the cloze procedure, and resolving the disagreements which arise concerning precisely what it measures, if we can answer a related question. Namely, what enables one and the same procedure to play two roles simultaneously for largely distinct populations, one as a language proficiency measure exclusively for second language learners, another as a reading comprehension measure primarily for native speakers? Tests of second language competence, for example, are not designed to discriminate levels of ability in both native and second language populations, yet cloze tests are apparently able to do so.[6] If we can specify what uniquely qualifies the cloze procedure to perform such a dual function, we may also have isolated those features which set it apart from more traditional measures.

The key to understanding how the cloze procedure manages to perform the function of both reading

comprehension and language proficiency measure simultaneously lies at the smallest unit of analysis possible for a cloze test - at the level of the word. To prefigure the account which follows, every cloze task will be seen to comprise two components: native speakers score consistently high on one (which we will call the "functional" component), but exhibit substantial variance on the other (which we will call the "content" component). Second language learners' cloze performance, in contrast, varies widely along both dimensions. A closer consideration of the literature will allow us to flesh out the features of these two components.

Researchers have captured this dichotomy by isolating the word types which go into making up cloze tests. Taylor (1956, 1957) found that certain types of words were very easy to replace, and other types relatively harder, for native speakers. His "easy words" were verb auxiliaries, conjunctions, pronouns, and articles, among the so-called "function" words in English. The "hard words" included nouns, verbs, and adverbs, traditionally labelled "content" words by grammarians.[7] Abundant research supports such a dichotomous classification of words on the basis of ease of prediction (e.g., Aborn et al. 1959; Coleman 1971; Coleman & Blumenfeld 1963; Fillenbaum et al. 1963; Gough 1983; Greene 1965b; Hanzeli 1979; Kibby 1980; Nicol & Miller 1959; Porter 1983; Rankin & Thomas 1980; Shanahan et al. 1982; Smith-Burke et al. 1978; Stubbs & Tucker 1974; Weaver & Bickley 1967).

Taylor also found that constructing cloze tests using exclusively "easy" as opposed to exclusively "hard" words as items led to different outcomes. When he attempted to measure information gain by comparing performance on a cloze test administered before reading an intact passage with scores on a cloze test administered afterwards, he found that the gains evidenced on the "hard" word cloze tests were significantly larger than on tests deleting only "easy" words (Taylor 1956, 1957). That is, when "content" words were the ones deleted, subjects were much more successful at replacing these after having read the intact passage - these items were presumably able to reflect learning that had taken place. When only "function" words were deleted, however, not only were scores on both pre- and post-passage cloze tests significantly higher than any achieved on the "content" cloze tests, but much less gain was evident (cf. Rankin 1959a, 1974).

This is precisely what we would expect in the case of native speakers (all college students, in these studies). The "easy" words, on the whole, were predictable based not on passage content, but rather by virtue of these readers' native competence in the language. These are the sorts of items we would not expect a native speaker to "learn" from reading a particular passage, but rather they should remain stable across administrations of the test. In contrast, prediction of the "hard" words required the gathering of sense from contexts larger than the clause or the sentence, relying on a more global understanding of the message and

reflecting knowledge specific to the passage content." Such items should be, and indeed were, more sensitive to information gain.

Rankin (1959a) hypothesized that when using an any-word deletion system, in which all types of words are subject to removal, two distinct realms of language ability should be tapped. The first would be competence in the structural redundancies of the English language that are relatively free from the influences of global context; the second would represent the knowledge of word meanings and semantic relationships that is brought to bear in comprehending the overall message of the text. While the "function" word / "content" word distinction could not serve to specify these two components exhaustively, since local redundancies can occur in semantics as well as in syntax (Rankin 1978), this dichotomy seems to be reflected in the dual nature of cloze items, nonetheless.

If we acknowledge these two components of cloze test measurement, it then becomes possible to explain how one test can serve both as a measure of reading comprehension and language proficiency for distinct populations. The cloze procedure takes on the characteristics of either type of test according to the ability which subjects bring to the task. For native speakers, functional deletions are "easy," that is, these subjects are uniformly successful in replacing such items based on their shared native competence in language structure. The contribution of this component to overall cloze scores varies little, and does not permit the

fine discrimination in ability which the cloze procedure is in fact able to make among native speakers. That variance comes, rather, from the content component, and accounts for the healthy correlations with readability and reading comprehension measures."

For learners of a second language, just as for native speakers, the content component is challenging, and so contributes a large share to the total variance. This population, however, is also striving to learn and master those functional elements of language which native speakers already command. For the second language learner, then, <u>both</u> components must be viewed as "hard," in the sense that they both will yield a wide range of scores depending on progress made in acquiring the second language (cf. Wainman 1979). Overall scores will continue to correlate with reading comprehension tests, except that variance will now spring from two primary sources rather than from one.[10]

That two such components exist, placing differential demands on native as opposed to second language speakers, is supported by a study conducted by Oller and Inal (1971). They constructed a cloze test by deleting only prepositions, a representative word class of the functional component, and administered this using acceptable alternate scoring to groups of native and second language university students. They found that the native speakers performed so uniformly well on the task (mean score 97%), that it was not possible to discriminate levels of ability. The second language students, on the other hand, showed a wide range of variance

in their scores, and were differentiated from native speakers on overall level of performance. This study makes it clear how second language performance on a cloze test could well exhibit variance from a source separate and distinct from that drawn on by native speakers.[11]

Conclusion. The cloze procedure is capable of measuring both language competence and reading comprehension, largely as a function of the abilities which subjects themselves bring to the task. The more a student is unfamiliar with the language in question, the more sensitive the cloze procedure becomes to discriminating levels of competence in the relatively local syntactic and semantic regularities of that language. This occurs because those cloze items which reflect such redundancies will contribute a relatively greater amount to the total score variance in these cases than they would otherwise. In short, the cloze procedure is able to take on the role of a language assessment device.

For native speakers, the cloze procedure functions quite well to discriminate levels of reading ability, since their performance is relatively unaffected by difficulties with linguistic structure per se. The cloze items which will generate the greatest amount of variance for this population will be those items which tap competence in such skills as understanding the sense of the passage, draving inferences across sentences, and predicting based on the content of the message.[12] The reading comprehension of second language learners can also be measured using cloze, but the further removed from native competence subjects are, the less

trustworthy cloze becomes as an uncontaminated measure of reading comprehension. That is, the inability to make certain word predictions due to plain unfamiliarity with many common redundancies of the language would introduce variability into the total score of these subjects that is not necessarily related to their ability to comprehend text.

6.2 A Perspective on the Literature

This model of cloze provides us with a framework for resolving some of the discrepancies which have arisen in the literature examining what the cloze procedure measures.

First of all, the experimental manipulations of deletion patterns revieved in section 5.3 above are subject to differing findings depending on the sorts of words that are included in the design. These studies, which all showed no effect for increasing context between cloze blanks, used any-word deletion - that is, all types of words were randomly subject to removal. The deleted words should thus reflect the statistical frequency of word types found in the passage, with function words representing anywhere from 44% to 63% of the total (Coleman 1971). Because function words are easier for native speakers to predict (see section 6.1 above), the frequency of correct responses on these items should be higher and more uniform than on the remaining content words. Since deletion studies only examine relatively small stretches of text, high performance on these items, which reflect local syntactic and semantic constraints, could well mask the sensitivity of the remaining cloze items to broader contextual influences (cf.

Bachman 1982).

Two studies support this interpretation. Fleming et al. (1973) showed that adding context up to 18 words on either side of the missing word did increase accurate prediction. This occurred, however, only when the words under consideration were specific types of nouns and verbs, representative types of the content component. Rankin and Thomas (1980) constructed cloze passages by deleting every 5th, 6th, 10th, or 15th word, with any type of word subject to removal. Overall, under the any-word condition, they found no effects due to increased context. Yet, when they focussed exclusively on the content words which had been deleted, the effects of increased context around cloze blanks was positive and significant. This led the authors to suggest:

> Conflicting findings in the literature concerning the effects of broader context on cloze test performance may well reflect the use of materials by different investigators which had different percentages of function words in the message. The greater proportion of function words in a given message, the smaller the influence of greater context for the test as a whole should be (1980, p.53).

The distinction between function and content words is relevant to the scrambling studies as well (Cziko 1983; Henk 1982). Bormuth (1964) strongly urged researchers for purposes of test reliability to administer all possible cloze forms of a passage (see section 3.3 above). This should be done in order to ensure a proper balance of cloze items, since the type and difficulty of deleted words can vary widely from one cloze form to another (cf. Meredith &

Vaughan 1978; Porter 1978). All the scrambling studies revieved in section 5.4 used only one form of the cloze passage, and there are no indications that any of the researchers examined cloze items to see what proportion of the deleted words were truly dependent on inter-sentence understanding (cf. Bridge & Winograd 1982; Chavez-Oller et al. 1985). The fact that about half of these studies show no intersentential effects might indicate no more than that half of the cloze forms constructed over these passages were more heavily weighted towards local redundancy. Treating one random cloze form as representative of all forms would tend to obscure the contribution of one component or the other. These comments are relevant to those studies which add intact text to the beginning of cloze passages as well.

The factor analyses are also subject to this caveat. Bachman (1982), the only factor analyst to study second language learners, made sure that items tapping intersentential cohesion were represented in the cloze protocol. He argues that without identifying such items and including them as variables, a given cloze form (generated by a random deletion procedure) may "...sample a larger proportion of clause-bound words, and therefore appear to be measuring only lower-level skills (1982, p. 66)."

6.3 Summary

The language ability that subjects bring into the cloze task largely determines the function which it will perform. For speakers who share native competence in the redundancies of the language, it will be those cloze items that reflect

content understanding, that is, largely the content words, that will contribute most to the total score variation. For second language speakers, those test items which are sensitive to proficiency in the structural redundancies of language, i.e., largely the function words, will also contribute variance, and so the cloze procedure is able to serve this population as a measure of both language ability and reading comprehension.

This view of the cloze procedure, wherein the abilities of the subject are paramount in determining how to interpret the results of the test, also allows us a developmental perspective on the cloze test which has not received attention in the literature. For native speakers at developmentally early stages, before a stable base level of linguistic competence has been established, cloze test results would reflect variation both in their command of linguistic redundancy and in their ability to comprehend the message of the text. In these cases, it is only when students happened to share roughly the same developmental level of grammatical ability that variance in cloze test results could reflect differences in content understanding primarily. As students developed competence in the redundancies of language and came to show more uniform performance on such items with increasing age, the cloze procedure would accordingly become a more accurate discriminator of content understanding.[13]

Since cloze items are differentially able to reflect grammatical as well as semantic abilities, the cloze

procedure can also serve as a diagnostic tool for isolating those cases where linguistic competence is for some reason delayed or divergent. A study by Willows and Ryan (1981) employed the cloze procedure in just such a capacity to examine the grammatical ability of native speaking readers in the fourth, fifth, and sixth grades. They found that, regardless of grade level, students who were identified as skilled readers were more successful in making grammatically correct replacements on a cloze test than were less skilled readers. Research by Cambourne (1977) and McDaniel (1983) yielded similar results. Such analyses confirm the capacity of the cloze procedure to uncover potential sources of reading difficulty that may not be apparent from performance on more traditional reading comprehension measures.

Having arrived at this perspective on the cloze procedure, we are now ready to consider the performance of deaf subjects on these tasks. Acknowledging that the English language ability of deaf students is often divergent in comparison to hearing subjects (for example, see Quigley & Paul 1984), and that reading comprehension ability lags behind as well (for example, see King & Quigley 1985), we stand alerted to the prospect that the cloze scores of this population may well reflect large variation in both functional and content components.

7.0 Deaf Performance on the Cloze Procedure

The literature concerning the performance of deaf students on the cloze procedure is sparse, but what studies there are coincide with several types of research design we

have already considered for hearing populations.

Certain investigations are relevant to issues related to cloze test format (Fischler 1983; LaSasso 1978b, 1979; LaSasso & Davey 1983a; Webster *et al.* 1981; Wilson *et al.* 1980; Wolk & Schildroth 1984; Wood *et al.* 1981). Other studies have examined the effects of varying deletion patterns (Odom *et al.* 1967) or of adding context around cloze blanks (Marshall 1970a,b). We will consider these three types of inquiry first, as a prelude to examining the practical validation studies.

The remaining research considers the ability of deaf cloze performance to rank text in the same manner as independent criteria do (LaSasso 1978b, 1980; Moores 1967, 1970), or seeks correlations between cloze scores and standard tests of reading comprehension (LaSasso 1978b; LaSasso & Davey 1983b). An examination of these studies will conclude the present review of literature, and bring us back to the focus of this qualifying paper. We will be in a position, after considering these final pieces of evidence, to draw conclusions concerning the role which the cloze procedure has to play as an assessment device for the deaf.

7.1 Test Format

Evidence that multiple-choice test items may not be the ideal means by which to assess the reading comprehension of the deaf comes from LaSasso's (1978b, 1979) finding that a certain type of question form, viz., the WH-question, is easier for deaf students to answer than are other sorts of multiple-choice items, regardless of content. This means

that deaf performance on multiple-choice measures could well show variance based not on an understanding of the text itself, but rather on the relative difficulty of certain question types. Researchers have also found that deaf students may rely on simple visual matching strategies (LaSasso & Davey 1983a; Webster et al. 1981; Wilson et al. 1980) or on inappropriate word associations (Wolk & Schildroth 1984; Wood et al. 1981), rather than on a genuine understanding of the text, to choose between multiple-choice alternatives. These results suggest that a method capable of assessing reading comprehension which does not have recourse to the multiple-choice format may be desirable in the case of the deaf. In theory, the cloze procedure would appear to be a prime candidate.

That the cloze procedure has the potential to serve as a measure of reading and language ability for the deaf is suggested by a study by Fischler (1983). He calculated the degree of agreement which hearing subjects evidenced when replacing words deleted from incomplete sentences. Those items that were highly constrained for the hearing, i.e., those word replacements on which hearing subjects largely agreed, were also the items on which deaf subjects showed the strongest consensus. It would appear from this study that cloze items are able to elicit similar patterns of response from hearing and deaf subjects.

7.2 Manipulation of Deletion Patterns

Odom et al. (1967) examined the effect of constructing cloze passages with every third, fourth, or fifth word

deleted. The cloze tests were administered to deaf high school students and it was found that they were better able to predict the form class of words (that is, to supply the correct part of speech, not necessarily the verbatim response) as the frequency of deletion decreased from every third to every fifth word. They concluded that deaf cloze performance is sensitive to varying amounts of contextual constraint.

These findings conform quite nicely with the results of deletion experiments with non-hearing-impaired populations (see section 5.3 above). Those studies found that deletion patterns more severe than every fifth word had progressively adverse effects on subject performance.

7.3 Manipulation of Larger Contexts

Marshall (1970a,b) investigated the performance of deaf high school students on cloze tasks with three levels of surrounding context: cloze items presented in the original full passage; the same items contained in isolated sentences presented in random order; and these items imbedded in sentence fragments only. A deletion scheme was used in which 20 content and 30 function words were removed whose form class was predictable even from the sentence fragments. He found that cloze performance was significantly more accurate in the context of the full passage than on the sentence fragments. His conclusion was that the cloze scores of deaf subjects are sensitive to contextual constraints beyond the clause and sentence level, a finding not out of line with a portion of the research on hearing cloze performance.

7.4 Ranking of Text

The research design of comparing passage ranking, as determined by mean cloze scores, with the ranking of the same passages by an independent criterion, has been employed with the deaf by two researchers (LaSasso 1978b, 1980; Moores 1967, 1970).

Moores' deaf subjects, whose reading level ranged from 4.0 to 6.0 on the paragraph meaning and word meaning sections of the Stanford Achievement Test (mean score of 4.77), took cloze tests constructed using stories from basal readers determined by the publishers to be appropriate for the fourth, sixth, and eighth grade levels. Analysis of variance revealed that deaf performance differed across levels of text difficulty. The highest performance was observed on the passage from the fourth grade text, as would be expected. However, while scores on the selections from the sixth and the eighth grade texts were substantially lower than on the fourth grade passage, they were almost identical to one another. It would seem from the magnitude of the scores (fourth grade score was nearly three times greater than either the sixth or the eighth grade score) that the cloze scores were only able to distinguish the simplest text from the two more difficult passages. In contrast, hearing students matched with the deaf students for reading level on the SAT ranked all three texts in order of difficulty.

In the second study, LaSasso (1978b, 1980) constructed cloze tests from narrative stories rated at the third,

fifth, and seventh grade levels by the Dale-Chall and Fry readability formulae. These were administered to a group of deaf high school students whose reading level ranged from 4.0 to 7.6 according to performance on the reading comprehension subtest of the Stanford Achievement Test (mean score of 5.9). LaSasso found dramatic divergences in expected ranking, which were statistically significant. The third grade passage was hardest for the students, the fifth grade passage easiest, and the seventh grade passage somewhere in the middle. When she broke down the grouping of subjects further into reading ability levels according to the SAT (4-4.9, 5-5.9, and 6-7.6), identical rankings of the three passages emerged for each group. LaSasso was not able to determine whether it was the cloze procedure, the readability formulae, or both, which accounted for these findings. Hearing controls were not employed in this phase of LaSasso's research.

Moores suggests a possible reason for the failure of the deaf subjects in his study to rank all three passages in order. Namely, it may have been the case that the two more difficult passages were beyond the competence of the deaf students, "...that the deaf group reached an asymptote at the sixth grade level that would [have been] maintained on passages at even the tenth or twelfth grade levels of difficulty...(1967, pp.96-97)." Clearly, an eighth grade passage may have been too difficult for a group of readers whose reading comprehension ranged only as high as the sixth grade level, and thus Moores' explanation seems plausible.

LaSasso's data, however, are more difficult to explain. The levels of text difficulty used in her study (third, fifth, and seventh grade) seem appropriately matched to the reading abilities of her subjects (4.0 - 7.6), yet passage ranking based on cloze scores deviates dramatically from expectations.

Conclusion and Suggestions for Research Design. On the whole, neither of these studies provides strong evidence that deaf performance on cloze tasks is able to rank text in the same way that independent criteria do. It is not clear from the research, however, whether it is the cloze procedure itself which accounts for these results.

The ranking study by LaSasso (1978b, 1980) lacks certain controls that would help clarify the role which the cloze procedure played in producing her surprising findings. Two design features for ranking studies are essential.

1) Hearing students of comparable reading ability, as determined by a separate measure, should always be included.

If both hearing and deaf perform in a similar manner, but diverge from expectations (such as ranking text passages differently than do independent ratings), then it would not be the specific interaction of deaf readers with cloze passages that could be implicated in these results. If the performance of the two groups diverges, with deaf performance somehow out of line and hearing performance in agreement with rankings, then additional evidence would be obtained for suggesting that cloze results may not be interpretable in the same way for deaf as they are for

hearing students. If the two groups agree in ranking, then the case for cloze will have been made that much stronger if they both follow expectations, or be seriously challenged if they both disagree with independent rankings. Hearing controls should be included in all correlational and criterion cloze score studies as well.

> 2) Readability formulae should not be the only independent ranking device used. In addition, the judgment of teachers of the deaf as to the relative difficulty of the test passages should be obtained.

Only passages on which the ranking of readability formulae and expert judgment agree should be included. In this way, if divergent results for the deaf are found, they could not be ascribed to peculiarities inherent in the readability formulae themselves.[24]

Studies of deaf performance need to articulate with second language research as well as with the research on hearing native speakers. We have seen how the cloze technique is able to become a different instrument (presumably taxing psycholinguistic processing mechanisms differentially) for second language learners than for native speakers. The question becomes, if it is to be used with deaf subjects, whether or not the demands placed by the task on their processing systems resemble more closely those placed on second language learners or on native speakers, or whether perhaps the cloze is essentially a third task for the deaf. Including second language subjects along with native speakers in the research design could yield comparative data that may well help answer this question.

Therefore,

> 3) Second language learners, matched for reading ability with deaf subjects and hearing controls, should also be included in every phase of the research, and the performance of all three groups compared.

Finally in this connection, it would be instructive to see if patterns of response could be found that would speak to the specific model of the cloze procedure put forward in this qualifying paper. If what we have called the functional and the content components were isolated for analysis, by deriving scores for specific word classes separately and computing their variance, it may be possible to detect differential performance on certain items by the deaf that would confirm that cloze is taking on the role of language proficiency device, in addition to reading comprehension measure, for this population.[15] In order to further explore this,

> 4) Researchers should not only describe how accuracy of response differs across categories of words, e.g., across certain function as opposed to certain content words; the variance which is generated by different word types should be subject to analysis as well.

7.5 Correlation with Traditional Tests

Two studies have examined the correlation of cloze scores with a standardized reading measure. LaSasso (1978b) found that deaf subjects' cloze performance, on the fifth grade passage used in the ranking study described above, correlated significantly but weakly (r=.47) with scores on the reading comprehension subtest of the SAT. LaSasso and Davey (1983b) computed a slightly higher correlation (r=.57) between cloze performance and this same subtest.[16]

A fifteen-question multiple-choice test was also constructed over the same fifth grade passage in the 1978 LaSasso study. Instead of comparing cloze scores with performance on the full set of multiple-choice items, however, correlations between cloze scores and each of three five-item subsets of these questions were computed separately (five questions probed text implicit and five questions text explicit information using incomplete statement stems; five questions were text explicit and employed WH-question stems). All three rho values were significant but exceptionally weak (.34, .17, and .21, respectively), due in great part to the small spread of scores (only 0 to 5) possible on each subtest. Had performance on the full multiple-choice test been compared with cloze scores, the resulting increase in variance might have yielded correlations of substantially greater magnitude.

Conclusion and Suggestions for Research Design. LaSasso's investigations do not yield convincing evidence for the validity of cloze testing as a measure of reading comprehension for the deaf. Neither do they, however, for the hearing. A group of hearing students, matched according to reading level with the deaf students in the 1978 study, showed low correlations between their cloze scores and all four reading comprehension measures that were indistinguishable from the deaf (for the hearing students, r=.49 between cloze and SAT performance, with rho values of only .38, .44, and .22 between cloze and same-passage performance).

The correlations in this study, for hearing as well as for deaf, are out of line with previous findings of research.

Correlations of cloze performance with multiple-choice measures constructed over the same passage have consistently been found higher than correlations with standardized tests for non-hearing-impaired subjects (see Rankin 1965, and section 4.2 above), not substantially lover as in LaSasso's study. Since it is the same-passage multiple-choice test which provides the basis for establishing the practical criterion cloze scores used to match readers with text in the classroom (see section 4.3 above), it is important that research design not preclude the possibility of finding meaningful correlations with such measures. To this end, two features should be incorporated into future investigations:

> 5] A variety of question types and an appropriate number of stimulus items should be used to construct the same-passage multiple-choice measure, in order to ensure that a large enough score spread is generated to maximize the possibility of obtaining meaningful correlations with cloze performance.
>
> 6] In addition, variance in cloze performance should be maximized by employing "acceptable alternate" scoring, in addition to verbatim scoring, for the deaf.

It will be recalled (see section 3.2 above) that investigators with second language students have often opted for use of acceptable alternate scoring due to the limited variance which verbatim scoring yielded for this population. The correlational studies described above used verbatim scoring only, and the lov rho values that were obtained are one of the possible consequences.

A gap which has heretofore existed in the correlational literature could also be filled if, as suggested in section 7.4, data from second language learners were to be collected.

Namely, of all the studies of second language cloze performance in relation to reading measures (see section 4.2), none has examined correlations of cloze scores with same-passage multiple-choice tests. Whether the pattern in evidence for native speaking subjects would hold up (same-passage correlations being higher than standardized-measure correlations), and whether the correlations would be just as strong, is as open a question for second language learners as it is for the deaf. Obtaining second language data in this phase of the research as well as in the next, the criterion cloze studies, could offer instructive comparisons with deaf performance at the most crucial points in the validation process.

Finally, traditional reading measures should not be the only independent standard considered for correlational analysis. There is no a priori reason to assume that deaf cloze performance might not correlate with measures of second language competence as well. One research study (Ragosta & Nelson 1986) found that a group of pre-college deaf students performed more like second language subjects than like native speakers on the TOEFL, suggesting the potential of this test as a language measure for the deaf. There is every possibility, then, that the cloze procedure might also be validated as a language assessment device for this population, as it has been for second language learners. Accordingly,

> 7) Correlations and comparisons between the cloze performance of the deaf and their performance on standard measures of second language proficiency should be sought.

7.6 Establishing Criterion Cloze Scores

No studies have attempted to establish criterion cloze scores for the deaf population. Several researchers (Cohen 1967; LaSasso 1978b; MacGinitie 1964; Marshall 1970b; Moores 1967; Odom *et al.* 1967), however, have found evidence which suggests that cloze performance may not predict multiple-choice scores for the deaf in the same way that it does for hearing students. Even when matched for reading ability on independent comprehension measures, the scores of deaf subjects in all six studies were consistently and significantly lower on word replacement tasks than were the scores of hearing subjects.[17] There is every reason to suspect, therefore, that the cloze scores achieved by a deaf student would not carry the same ready interpretation for a reading teacher as cloze scores that were earned by the hearing.

Conclusion and Suggestions for Research Design. In light of all this, the use of the cloze procedure to match deaf readers with text (LaSasso 1978a) would seem to be a very risky enterprise indeed. Not only have the investigations described in previous sections failed to marshal strong support for the use of the cloze procedure to assess the reading comprehension of the deaf, but the practical use of this technique for matching reader with text in the classroom would appear to rest on no empirical foundation whatsoever.

Research should venture beyond the studies to date and attempt to establish criterion cloze scores for deaf performance which are interpretable and useful for the

classroom teacher. If significant and substantial correlations are found between cloze scores and scores on same-passage multiple-choice tests, then

> 8) Multiple-choice scores should be regressed on verbatim cloze scores, and the resulting equation used to compute the cloze score equivalents of 75% and 90% correct on the multiple-choice measures, the traditional instructional range.

These resulting scores should then be compared to the widely accepted cloze criterion scores of 44% and 57% respectively, as well as to the performance of hearing controls. Since past research has shown deaf cloze scores to be significantly lower than hearing cloze scores even for "comparable" readers, a new set of criterion scores for the deaf could conceivably result from this analysis.[18]

> 9) Multiple-choice scores should be regressed also on "acceptable alternate" cloze scores, and the resulting equation used to compute the cloze score equivalents of 75% and 90% correct on the multiple-choice measures.

These resulting scores should then be compared to the traditionally accepted criterion cloze scores for the instructional range, as well as to the performance of hearing controls under <u>verbatim</u> scoring. Since the employment of acceptable alternate scoring would result in the overall raising of deaf cloze scores in comparison to verbatim scoring, the criterion scores generated by this method might more closely align themselves with the traditional hearing criteria.[19] If so, it would indicate that teachers of the deaf may be able to employ the standard and widely used criterion cloze scores in the classroom, under the condition that acceptable alternate scoring be used.

7.7 Summary

Theory suggests that the cloze procedure should be an appropriate measure of reading comprehension for the deaf. Deaf subjects appear to respond to cloze items in a manner similar to hearing subjects, and seem sensitive to changes in context when supplying responses. Research, however, has not yet validated the cloze procedure as a comprehension measure for this population in two senses: it has not shown cloze performance to reflect text difficulty in predictable ways, and correlations with standard reading measures are weak. No attempt has been made at all to establish criterion cloze scores that could define the instructional level of text.

8.0 Summary of Findings

This review of the literature has provided ample evidence that the cloze procedure is valid for use with both native speaking and second language populations. For second language learners, the cloze procedure has been studied as a means by which to assess their proficiency in the target language. For native speakers, it has served as a measure of reading comprehension. The cloze procedure is able to perform this dual function by virtue of the items which make up such tests - some are differentially sensitive to local constraints and redundancies, others are open to the broader influences of content understanding. That both components exist side by side in one and the same cloze test was made clear during our examination of the debate over what cloze tasks measure.

We gained a perspective on the cloze procedure which views the skills subjects bring to the task as crucial to

determining the role cloze will play. Those learning a second language will respond differently to certain items than do native speakers, thus allowing the cloze procedure to function as a language proficiency device for them. For native speakers, sharing a common language base, cloze results represent more directly their ability to derive meaning from text. In the case of those caught in the middle, as it were, subjects with linguistic deficiencies that interfere with successful reading, the cloze procedure is able to go beyond traditional reading tests to diagnose the presence of such difficulties. It is able to do so by virtue of its capacity to reflect not only content understanding, but competence in the structural redundancies of the language as well.

The cloze procedure would seem in theory to be an excellent way by which to measure both the language and the reading abilities of the deaf, yet we found that validation of cloze tests for reading assessment has not proven successful. One reason for this may lie in the quality of target language competence which the deaf themselves bring to the task. Exploring what the cloze procedure is able to reveal concerning the English language ability of the deaf may help to clarify the role that cloze has to play as a measure of reading comprehension for these subjects. Another reason that the cloze procedure has not yet been validated as a reading measure for this population, however, is the paucity of validation research in the field. Until more data are available concerning deaf performance on cloze tasks, data generated by studies incorporating research designs of the

type suggested in this qualifying paper, formal claims to validity cannot be made.

Finally, we were in a position to return to the question posed at the beginning of this qualifying paper: to what extent can teachers of the deaf make use of the cloze procedure as a practical technique by which to match readers with text in their classrooms? Our answer, in light of the research evidence to date, was that the basis for using the cloze procedure in such a capacity does not yet exist. While the cloze procedure may have potential for such application, the evidence necessary to establish confidence in this use is simply lacking.

9.0 Implications for Deaf Education

Teachers of the deaf are advised to avoid any use of the cloze procedure based on the traditional criterion scores of 44% and 57%. All evidence indicates that the cloze scores of deaf students do not index levels of reading ability in the same manner that they do for hearing native speakers.

The use of such scores to match deaf readers with text could, in fact, seriously compromise the quality of reading material selected for the curriculum. One of the few robust findings to emerge from the research on deaf cloze performance is that for a given level of reading ability, as determined by standardized reading tests, deaf students consistently underperform their hearing counterparts on cloze tasks. This means that decisions concerning text assignment, if they are made on the basis of the criterion cloze scores established for hearing students, would be likely to overestimate the

difficulty of text for a given deaf reader. For example, one might judge that a cloze score of 20% achieved by a deaf student would indicate a text too frustrating for that reader. Yet that score might well correspond to 75% comprehension on a multiple-choice measure, or otherwise indicate that the text was at a suitable level of difficulty. Assigning text only when a deaf student had met the hearing criterion cloze scores could result in a diet of material that was too simple and lacking in challenge (cf. Kelly & Ewoldt 1984).

The cloze procedure has not yet been validated in parametric terms for use with the deaf population. This does not mean, however, that teachers of the deaf should discount its usefulness as an assessment device. The diagnostic potential of the cloze procedure has been recognized since the early days of cloze research (e.g., Jenkinson 1957; Rankin 1959b). And two studies, examining its use specifically with deaf readers (Ewoldt 1981; Kelly & Ewoldt 1984), reveal that the cloze procedure may have remarkable utility for diagnosing the reading difficulties of this population, especially when innovative scoring procedures are employed.[20] The administration of cloze tests to readers on an individual, case by case basis, with the goal of isolating specific aspects of the reading task which students can and cannot negotiate successfully, remains a viable and potentially valuable tool for the classroom teacher regardless of whether large-scale validation studies of the cloze procedure prove to be successful.

In connection with the use of cloze as a diagnostic tool,

one consideration which teachers of the deaf should be aware of is the frustration that even hearing students can experience when faced with cloze tests administered for assessment purposes (cf. Almeida 1975; Cecil 1985; Hanzeli 1977; Rankin 1977). When lack of motivation sets in, no device, regardless of how sound it may be theoretically, can hope to assess a student's abilities accurately. I have personally observed underachieving deaf readers balk at cloze tasks, and have witnessed them give up in frustration when such tests were administered without feedback or prior exposure.

I have also used cloze passages for <u>instruction</u> with these same students, and have found them to respond with genuine interest and enthusiasm. The teacher can build positive feelings for the cloze procedure by using it, not only for the diagnosis of reading difficulties, but as an instructional device in the classroom as well. The cloze procedure seems to have a sort of Jekyll and Hyde personality - it can be engaging and stimulating when used for instruction, but daunting and frustrating when used for assessment. The teacher, by familiarizing students with cloze format in a positive context and using it to build confidence and skill in reading and language tasks, can thereby help maximize the accuracy of the cloze procedure when it is employed as a diagnostic technique.

There is no lack of suggestions available to teachers regarding instructional applications of the cloze procedure with both first and second language students (e.g., Berrent

1988; Bloomer 1962; Bortnick & Lopardo 1973; Carer et al. 1989; Devitz et al. 1987; Giordano 1985; Kennedy & Weener 1973; Meyer & Tetrault 1986; NY City Board of Education 1982; Peters 1983; Rankin 1959b, 1977; Riley 1986; Sampson et al. 1982; Valmont 1983; and see critical reviews by Jongsma 1971, 1980). One instructor of the deaf (LoMaglio 1986), in fact, has reported that using cloze exercises with his students was effective in developing critical thinking skills and language competencies important to reading success. Teachers of the deaf should thus, at the present time, consider using cloze tasks in their classrooms for two purposes: for instruction, and for diagnostic assessment.

10.0 Directions for Future Research

Research into the cloze procedure and its use with the deaf population should focus on three areas:

1) Validation studies should be pursued, with the specific goal of verifying cloze criterion scores. If particular scores are useful for matching reader with text, then teachers should know this so that they may make use of them in the classroom. If it is not possible to establish criterion scores for deaf students, then teachers need to know that as well so that they do not misuse this device.

2) Research should investigate further the potential of cloze to diagnose individual reading difficulties. Its most important application as an assessment device for the deaf population may only be describable in other than parametric terms.

3) Experimental studies need to examine interventions that unite cloze qua instruction with cloze qua assessment in single classrooms. It is in the union of instruction and assessment that cloze may offer its greatest potential benefit for students, hearing or deaf.

NOTES

1. Prior to Taylor's experiments and his coinage of the term "cloze", variations in text completion techniques were most often used for psychological testing purposes. For accounts of these earlier uses see Harris (1985) and Wardell (1985), as well as the comprehensive bibliography in Buros (1978).

2. Several other types of cloze format have been developed for use with both native and second language subjects, but it will be impossible to describe them all here. For in-depth treatments of some of the more well-known alternative formats, the interested reader may consult representative studies on matching cloze (Baldauf & Propst 1978, 1979; Baldauf et al. 1980; Cunningham & Cunningham 1978; Propst & Baldauf 1979); multiple-choice cloze (Hinofotis & Snow 1980; Jonz 1976; Scholz & Scholz 1981) including the maze procedure (Bradley & Meredith 1978; Bradley et al. 1978; Guthrie 1973); cloze-elide (Manning 1986, 1987; Mullen 1979a); and the C-test (Raatz 1985; Raatz & Klein-Braley 1981).

3. For details on other scoring systems based on degree of match with native speaker responses, the reader is referred to Carroll et al. (1959), and to studies on clozentropy (Darnell 1968) and the New Cloze Test (Hisama 1977; Hisama et al. 1977). McLeod and Anderson (1966, 1970) have also suggested a scoring system for native speakers based on degree of match with skilled reader responses.

4. Two other studies (Ransom 1966, 1968; Jones & Pikulski 1974) arbitrarily set cloze criteria for the instructional range (30% - 50%) based on a simple visual inspection of scores.

5. One study (Briere & Hinofotis 1979) experimented with various cloze scores as cutoff points for ESL placement, but did not seek to establish equivalent scores on any standard second language measure.

6. When one considers what is expected from items on a traditional test of foreign language proficiency (for example, see Clark 1977), it is not at all clear how one test could be viable for both a native and a second language population. Such tests should be able, at a minimum, to discriminate between native and non-native speakers. More precisely (Alderson 1980; Angoff & Sharon 1971), 1) native speakers with normal competence in the language should perform at uniformly high levels, with so little variance in performance that discrimination between skill levels is precluded; 2) non-native speakers, on the other hand, should show wide variance in scores according to their experience with the language, and the test should be able to rank them according to skill level.

The cloze procedure meets the second specification. Namely, performance of second language learners shows wide variance on cloze tests, and a large portion of this variance

is shared in common with other, standardized measures of foreign language proficiency (see section 4.2). Furthermore, cloze tests are sensitive enough to rank students according to their level of academic placement in second language programs (Bachman 1985; Briere et al. 1978; Briere & Hinofotis 1979; Hanzeli 1977; Heilenman 1983; Markham 1988; Oller & Conrad 1971; Porter 1978).

In regard to the first specification, however, the cloze procedure fails. It is not only second language learners who exhibit wide variance in performance on cloze tasks, but native speakers as well; enough variance, in fact, so that cloze performance (both of first and second language speakers) can be correlated with measures of reading comprehension. A test of second language ability on which native speakers do not top out carries such suspect validity for some researchers (for example, see Alderson 1980) that this becomes prima facie grounds for questioning the test.

7. A taxonomy of English word types often used by cloze researchers was presented by Fries (1952). He found that fifteen groups of "function" words (e.g., determiners, modal auxiliaries, prepositions) make up about 1/3 of the word tokens in text, but comprise only slightly more than 150 separate items. Four classes of words (nouns, verbs, adjectives, certain adverbs) were found to represent about 2/3 of the word tokens, and to include in their number thousands of individual items.

8. Research with native speakers by Carroll et al. (1959) also supports such a dichotomy of cloze items, at least for certain word types. In an experiment employing the scrambling format described in section 5.4, these researchers found that verbatim prediction of prepositions was stable across both coherent and scrambled conditions. That is, such items seemed insensitive to overall passage content. The prediction of nouns and verbs, in contrast, suffered under the scrambled condition. Carroll et al. remark that these results are consistent with the thesis "...that 'content-bearing' words such as nouns and verbs are more likely to be influenced by paragraph cues than words which are more concerned with the syntactical skeleton, so to speak, of the message (1959, pp. 61-62)."

9. Performance of native speakers on second language measures, in contrast, would not correlate with cloze scores. Overall variance on the cloze task would be large, reflecting performance on the content component, but overall variance on the second language measure quite small, as a result of these speakers sharing native linguistic competence.

10. Scores will also correlate with tests of second language proficiency, because such measures are especially sensitive to variance in the functional component.

11. It is difficult to find studies which speak directly to

the thesis that first and second language speakers will exhibit distinctive patterns of performance variance on cloze tests of traditional design. Of all the studies contrasting performance on function as opposed to content words (see section 6.1), only two (Porter 1983; Shanahan et al. 1982) report standard deviations. Both studies show, for native speaking college students, that variance in performance differs little between the two types of words when verbatim scoring is used.

Acceptable alternate scoring was also employed by Porter (1983), however, and the resulting pattern of variance was dramatically different. Using this scoring method, variance for native speakers was substantially less for function than for content word replacements. This suggests that the scoring system employed may be a critical design variable for isolating the contribution of the two components.

Second language speakers were included in Porter's study as well, and the comparison with native performance is most revealing. Unlike native speakers, variance in the performance of second language students on function words was actually higher than for content words under both scoring systems. For these two populations, then, total cloze score would reflect variance contributed differentially by two distinct sources.

12. This model of the cloze test as simultaneously a reading and a language proficiency measure is really not all that different in structure from tests of second language competence, when one looks more closely at the make-up of such tests. Contrary to Alderson's (1980) claims, native readers need not necessarily top out on all components of such tests. For example, Clark (1977) found that among the subtests of the TOEFL, error rates for native speakers were large on only one measure, the Reading Comprehension section, a test modeled on the traditional multiple-choice paradigm. He reports that

> ...percent-fail levels for the native English speakers were...high..., suggesting substantial difficulty on the part of the examinees in responding correctly to certain of the reading questions. Closer analysis indicated that a high proportion of these items dealt with the summarization or interpretation of passage content, although several of the more factually oriented questions also proved difficult for the examinee group (p.29).

Native speaker performance thus shows large variance on one specific portion of the foreign language test, while evidencing little variance on others (cf. Angoff & Sharon 1971). For second language speakers, in contrast, performance varies across every portion of the test. So it is with cloze tasks, though these components are not to be located, as in the case of the TOEFL, in isolated subsections of the test, but rather in the test items (deleted words) themselves.

13. Evidence from a study by Willows and Ryan (1986) supports the view that young native speakers steadily develop and then plateau in their ability to use local redundancies to replace words missing in text. They found that first, second, and third grade students made a significantly greater proportion of grammatically appropriate replacements at each succeeding grade level. In a similar study of grades four, five, and six (Willows & Ryan 1981) the ability to use local redundancies in this way had flattened out, with no increases apparent between grades. Skilled readers did make more successful use of contextual cues than less skilled readers, regardless of grade level, but by the higher grades even the less skilled readers were supplying grammatically acceptable responses in an average of 92% of their attempts (1981, p. 611).

14. In a follow-up study, LaSasso (1982) asked experienced teachers of the deaf to rate the passages used in the ranking study described above for difficulty. Teachers' judgments matched the readability rankings exactly, thus lending credence to the possibility that it was not the readability formulae but rather the interaction of deaf readers with the cloze task that had accounted for the aberrant results in her original ranking study.

LaSasso (1987) undertook still another investigation into the difficulty of these same passages, the results of which pointed this time in the opposite direction, that is, towards the readability formulae as being inaccurate. In this second follow-up, she collected ratings only from experienced teachers of the deaf who were themselves hearing-impaired (in the 1982 study, all but two of the 74 teachers had been hearing). These new ratings did <u>not</u> match the original readability rankings, and in fact came close to predicting deaf cloze performance. These results, coupled with the fact that these hearing-impaired teachers isolated certain factors as major contributors to text difficulty that were not emphasized by the predominantly hearing group or measured by the readability formulae, led LaSasso to conclude that investigators in the future should make sure to consult the "...judgments of teachers who are themselves hearing-impaired and [who] may have valuable insight into elements which contribute to text difficulty for hearing-impaired readers (1987, p. 402)." It should be noted that, along with being hearing-impaired, the second group of teachers also had more than 50% more teaching experience on the average than the 1982 group (mean years of experience = 11.39 compared to 7.2 years). Thus, amount of teaching experience as well as hearing status may have accounted for the apparently greater acuity of teacher judgments in the second group.

One conclusion which may be drawn from these two studies is that the judgment of only the most qualified group of experts possible should be sought for the purposes of ranking text. Unfortunately, the combined results of these two post hoc treatments do little to illuminate the role which cloze had to play in all this. Accepting the premise that the 1987 teachers, on the whole, were the more qualified group, and

given the fact that the difficulty ratings of this group did not match the readability rankings, this qualifying paper is suggesting that the original passages should probably not have been selected for the ranking study in the first place. It has been established that readability formulae can be "fooled" by outwardly simple prose that otherwise contains conceptually difficult elements, while cloze scores, in contrast, are able to reflect this greater difficulty (Taylor 1953). Perhaps this was the case in LaSasso's original study, that the cloze performance of deaf students was in fact a better gauge of passage difficulty for these subjects than were either the readability formulae or the hearing teachers' rankings. The point is that when the purpose of research is explicitly to examine the validity of the cloze procedure, then passages must be selected which will not set up road blocks to arriving at clear and cogent conclusions concerning the very object of the study. Research design that requires agreement between readability formulae and expert judgment (of the most qualified raters possible) would help to remove such obstacles.

15. Some investigators have sought to describe patterns in word replacements made by the deaf. Such inquiries have not been successful to date, however, in painting a consistent picture of deaf cloze performance.

Moores comments (1967, p. 102) that the deaf group and the hearing controls performed quite differently on various individual word replacements, but he does not, with the exception of the determiner "the" (1967, pp. 96-97), submit the response patterns of either group to detailed analysis. He urges further study in this area and stresses the potential of cloze item analysis for the diagnosis and remediation of language difficulties.

Marshall (1970a,b) analyzed the error substitutions of his hearing and deaf subjects according to word type, but concludes:

> No patterns of symmetry emerged such as nouns being substituted for pronouns or auxiliary verbs for main verbs, and vice versa. Instead there appeared to be a general scatter among the intrusions, extending across all levels of context...
>
> Similar conclusions were drawn for the function word intrusions ... (1970a, p. 688)

Marshall did report that both hearing and deaf performed better overall on function as opposed to content word replacements, a finding identical to that of MacGinitie (1964). Hart and Rosenstein (1964), on the other hand, found that deaf students reversed the normal hearing pattern, by evidencing greater difficulty in replacing function than content words. Still another researcher (Cohen 1967) reported only that hearing and deaf showed similar patterns of performance.

Such a confusing and sometimes contradictory array of findings would seem to support Moores' suggestion that response analysis may be more appropriate for diagnosing and remediating individual difficulties than for distinguishing populations parametrically. Nevertheless, this qualifying paper is suggesting that an analysis which focusses on describing patterns of performance variance across word types is worth pursuing, and may have a contribution to make to the theoretical understanding of the role cloze has to play as an assessment device for the deaf population.

16. Kelly and Evoldt (1984) examined the relationship between the cloze scores of deaf students and their performance on the reading subsection of the SAT in a different manner. They found that the two measures agreed to a statistically significant extent in their identification of minimally competent students. The authors did not report the exact scores which were deemed minimally acceptable on either of the two measures, however, nor did they compute a correlation coefficient for performance on the two tests.

17. The specific purpose for using cloze tests in Moores' (1967, 1970) investigation was, in fact, to uncover syntactic and semantic deficiencies in deaf cloze performance which are not reflected by their scores on standardized reading tests. That deaf cloze scores fell below those of hearing controls in all six of the studies comparing the two groups supports Moores' assertion (1967, pp. 59-60; 1970, p. 646), and a major thesis of this paper, that cloze is sensitive to variance in linguistic ability which is essentially untapped by traditional measures of reading comprehension.

I must disagree, however, with the associated claim made by Moores (1967, pp. 41, 95; 1970, p. 650) that lower cloze performance by the deaf indicates that standardized reading tests actually <u>overestimate</u> their reading ability. While the skills needed to supply precise replacements for missing words may be lacking, resulting in low cloze scores, separate and distinct strategies may nonetheless be applied by deaf students when reading a passage that can result in successful comprehension of the message. Despite the finding that some deaf may have recourse to inappropriate strategies to answer multiple-choice questions (see section 7.1), the resourceful deaf reader may also be able to rally common sense, use inference, apply prior knowledge, reason through semantic relationships, and otherwise utilize all manner of appropriate strategies to come to a genuine understanding of a text. The paucity of linguistic skill revealed by cloze makes the fact that deaf readers outperform expectations on reading comprehension measures that much more remarkable. Such accomplishments should not be shortchanged solely on the basis of poor cloze performance.

It should be noted that since the completion of his dissertation, Moores has come to hold a different viewpoint concerning standardized tests and what they reveal about the reading ability of the deaf. Moores reports (1987a; 1987b, pp.

285-286) that evidence is accumulating to the effect that standardized tests may, in actuality, underestimate the true reading ability of the deaf and their capacity to function in a literate society.

18. In one study (Kelly & Evoldt 1984) only 6% of deaf students tested with cloze on grade level material were able to meet the minimum 44% correct, using verbatim scoring. This suggests that the Bormuth criteria may be too stringent for use with deaf students.

19. Kelly and Evoldt (1984) also found that use of acceptable alternate scoring resulted in substantially more deaf students achieving the minimum 44% correct over grade level material (43% of students under acceptable alternate scoring as opposed to only 6% of students using verbatim scoring). Clearly, acceptable alternate scoring had the effect of raising deaf students' cloze scores, bringing them more closely into line with hearing standards.

20. For example, by applying two scoring criteria to the cloze responses of deaf subjects Randy and Monica, viz., "judged meaningful in the passage" and "judged meaningful in the sentence," in addition to scoring verbatim reproductions, Kelly and Evoldt (1984) were able to demonstrate very useful insights into the differential reading abilities of the two students. Randy appeared to be a far inferior reader compared to Monica based on overall verbatim responses (14% correct as opposed to 41% correct), but this information alone could not provide any clue as to the source of his difficulties. Responses judged meaningful only in the sentence (not, simultaneously, in the passage), however, were found to be more than twice as prevalent for Randy (23%) as for Monica (10%). The authors used this information to propose an instructional intervention for Randy designed to help him expand his view beyond the narrow limits of sentence boundaries when attempting making sense of a reading passage.

REFERENCES

Aborn, M., Rubenstein, H., & Sterling, T. D. (1959). Sources of contextual constraint upon words in sentences. Journal of Experimental Psychology, 57, 171-180.

Aitken, K. G. (1977). Using cloze procedure as an overall language proficiency test. TESOL Quarterly, 11, 59-67.

Alderson, J. C. (1979a). Scoring procedures for use on cloze tests. (ERIC Document Reproduction Service No. 208 683)

Alderson, J. C. (1979b). The cloze procedure and proficiency in English as a foreign language. TESOL Quarterly, 13, 219-227.

Alderson, J. C. (1979c). The effect on the cloze test of changes in deletion frequency. Journal of Research in Reading. 2, 108-119.

Alderson, J. C. (1980). Native and nonnative speaker performance on cloze tests. Language Learning, 30, 59-76.

Alexander, H. W. (1969). Investigation of the cloze procedure as a measuring device designed to identify the independent, instruction, and frustration reading levels of pupils in the intermediate grades. Dissertation Abstracts, 29/12-A, 4314-4315. (University Microfilm No. 69-10625)

Almeida, P. M. (1975). Practical implications of the cloze test for placement in instructional materials. Unpublished doctoral dissertation, Harvard Graduate School of Education.

Anderson, J. (1971). Selecting a suitable "reader": Procedures for teachers to assess language difficulties. RELC Journal, 2, 35-42.

Anderson, J. (1972). The application of cloze procedure to English learned as a foreign language in Papua and New Guinea. English Language Teaching, 28, 66-72.

Angoff, W. H., & Sharon A. T. (1971). A comparison of scores earned on a test of English as a foreign language by native American college students and foreign applicants to U.S. colleges. TESOL Quarterly, 5, 129-136.

Bachman, L. F. (1982). The trait structure of cloze test scores. TESOL Quarterly, 16, 61-69.

Bachman, L. F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. TESOL Quarterly, 19, 535-556.

Baldauf, R. B., Dawson, R. T., Prior, J., & Propst, I. K. (1980). Can matching cloze be used with secondary ESL pupils? Journal of Reading, 23, 435-440.

Baldauf, R. B., & Propst, I. K. (1978). Preliminary evidence regarding the validity of a modified cloze procedure for lower elementary ESL students. Educational and Psychological Measurement, 38, 451-455.

Baldauf, R. B., & Propst, I. K. (1979). Measuring ESL reading achievement with matching cloze. (ERIC Document Reproduction Service No. 186 468)

Bartoo, R. K. (1975). The effect of cumulative context upon cloze task performance of tenth graders. Dissertation Abstracts International, 36/02-A, 842. (University Microfilm No. 75-16951)

Berrent, H. I. (1988). Using cloze for reading comprehension. Academic Therapy, 23, 383-387.

Betts, E. A. (1946). Foundations of reading instruction. New York: American Book Company.

Bloomer, R. H. (1962). The cloze procedure as a remedial reading exercise. Journal of Developmental Reading, 5, 173-181.

Bormuth, J. R. (1963a). Cloze as a measure of readability. In J. A. Figurel (Ed.), Reading as an intellectual activity. International Reading Association Conference Proceedings, Volume 8. New York: Scholastic Magazines.

Bormuth, J. R. (1963b). Cloze tests as measures of readability and comprehension ability. Dissertation Abstracts, 23/11, 4218-4219. (University Microfilm No. 63-2586)

Bormuth, J. R. (1964). Experimental applications of cloze tests. In J. A. Figurel (Ed.), Improvement of reading through classroom practice. International Reading Association Conference Proceedings, Volume 9. Newark, Delavare: IRA.

Bormuth, J. R. (1965a). Optimum sample size and cloze test length in readability measurement. Journal of Educational Measurement, 2, 111-116.

Bormuth, J. R. (1965b). Validities of grammatical and semantic classifications of cloze test scores. In J. A. Figurel (Ed.), Reading and inquiry. International Reading Association Conference Proceedings, Volume 10. Newark, Delavare: IRA.

Bormuth, J. R. (1967). Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 10, 291-299.

Bormuth, J. R. (1968). Cloze test readability: Criterion reference scores. Journal of Educational Measurement, 5, 189-196.

Bormuth, J. R. (1969a). Empirical determination of the instructional reading level. In J. A. Figurel (Ed.), Reading and realism. Proceedings of the Thirteenth Annual Convention, International Reading Association, Part 1. Newark, Delaware: IRA.

Bormuth, J. R. (1969b). Factor validity of cloze tests as measures of reading comprehension ability. Reading Research Quarterly, 4, 358-365.

Bormuth, J. R. (1975). The cloze procedure: Literacy in the classroom. In W. D. Page (Ed.), Help for the reading teacher: New directions in research. National Conference on Research in English Bulletin Series, ERIC Clearinghouse on Reading and Communication Skills. Washington, DC: National Institute of Education.

Bortnick, R., & Lopardo, G. S. (1973). An instructional application of the cloze procedure. Journal of Reading, 16, 296-300.

Bradley, J. M., Ackerson, G., & Ames, W. S. (1978). The reliability of the Maze procedure. Journal of Reading Behavior, 10, 291-296.

Bradley, J. M., & Meredith, K. (1978). The reliability of the Maze procedure for intermediate and junior high school students. In P. D. Pearson & J. Hansen (Eds.), Reading: Disciplined inquiry in process and practice. Twenty-seventh Yearbook of the National Reading Conference. Clemson, South Carolina: NRC.

Bridge, C. A., & Winograd, P. N. (1982). Readers' awareness of cohesive relationships during cloze comprehension. Journal of Reading Behavior, 14, 299-312.

Briere, E. J., Clausing, G., Senko, D., & Purcell, E. (1978). A look at cloze testing across languages and levels. The Modern Language Journal, 62, 23-26.

Briere, E. J., & Hinofotis, F. B. (1979). Cloze test cutoff points for placing students in ESL classes. In E. J. Briere & F. B. Hinofotis (Eds.), Concepts in language testing: Some recent studies. Washington, DC: TESOL.

Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. The Modern Language Journal, 64, 311-317.

Brown, J. D. (1983). A cloze is a cloze is a cloze? In J. Handscombe (Ed.), On TESOL '83, The question of control. Selected papers from the annual convention of Teachers of English to Speakers of Other Languages. (ERIC Document Reproduction Service No. 275 137)

Buros, O. K. (Ed.) (1978). Cloze procedure (Ebbinghaus completion method) as applied to reading. In The eighth mental measurements yearbook, Volume II, 1163-1178. Highland Park, New Jersey: Gryphon Press.

Cambourne, B. (1977). Some psycholinguistic dimensions of the silent reading process, A pilot study. (ERIC Document Reproduction Service No. 165 087)

Carr, E., Devitz, P., & Patberg, J. (1989). Using cloze for inference training with expository text. The Reading Teacher, 42, 380-385.

Carroll, J. B. (1972). Defining language comprehension: Some speculations. In R. O. Freedle & J. B. Carroll (Eds.), Language comprehension and the acquisition of knowledge. Washington, D.C.: Winston and Sons.

Carroll, J. B., Carton, A. S., & Wilds, C. (1959). An investigation of "cloze" items in the measurement of achievement in foreign languages. Cambridge, Massachusetts: Harvard Graduate School of Education.

Carver, R. P. (1970). What is reading comprehension and how should it be measured? In G. B. Schick & M. M. May (Eds.), Reading: Process and pedagogy. Nineteenth Yearbook of the National Reading Conference, Volume 1: Process. Milwaukee, Wisconsin: NRC.

Carver, R. P., Johnson, R. L., & Friedman, H. L. (1971-72). Factor analysis of the ability to comprehend time-compressed speech. Journal of Reading Behavior, 4, 40-49.

Caulfield, J., & Smith, W. C. (1981). The reduced redundancy test and the cloze procedure as measures of global language proficiency. The Modern Language Journal, 65, 54-58.

Cecil, N. L. (1985). Instructional cloze: Confronting some common concerns. Reading Horizons, 25, 95-97.

Chavez-Oller, M. A., Chihara, T., Weaver, K. A., & Oller, J. W. (1985). When are cloze items sensitive to constraints across sentences? Language Learning, 35, 181-205.

Chihara, T., Oller, J., Weaver, K., & Chavez-Oller, M. A. (1977). Are cloze items sensitive to constraints across sentences? Language Learning, 27, 63-69.

Clark, J. L. D. (1977). The performance of native speakers of English on the test of English as a Foreign Language. TOEFL Research Reports, No. 1. Princeton, New Jersey: Educational Testing Service.

Cohen, S. R. (1967). Predictability of deaf and hearing story paraphrases. Journal of Verbal Learning and Verbal Behavior, 6, 916-921.

Coleman, E. B. (1971). Developing a technology of written instruction: Some determiners of the complexity of prose. In E. Z. Rothkopf & P. E. Johnson (Eds.), Verbal learning research and the technology of written instruction. New York: Teachers College Press.

Coleman, E. B., & Blumenfeld, J. P. (1963). Cloze scores of nominalizations and their grammatical transformations using active verbs. Psychological Reports, 13, 65-654.

Cunningham, J. W., & Cunningham, P. M. (1978). Validating a limited-cloze procedure. Journal of Reading Behavior, 10, 211-213.

Cziko, G. A. (1978). Differences in first and second language reading: The use of syntactic, semantic and discourse constraints. The Canadian Modern Language Review, 34, 473-489.

Cziko, G. A. (1983). Another response to Shanahan, Kamil, and Tobin: Further reasons to keep the cloze case open. Reading Research Quarterly, 18, 361-364.

Dale, E., & Chall, J. S. (1948). A formula for predicting readability: Instructions. Educational Research Bulletin, 28, 37-54.

Darnell, D. K. (1963). The relationship between sentence order and comprehension. Speech Monographs, 30, 97-100.

Darnell, D. K. (1968). The development of an English language proficiency test of foreign students using a clozentropy procedure: Final report. (ERIC Document Reproduction Service No. 024 039)

Dewitz, P., Carr, E. M., & Patberg, J. P. (1987). Effects of inference training on comprehension and comprehension monitoring. Reading Research Quarterly, 22, 99-121.

Entin, E. B., & Klare, G. R. (1978). Some inter-relationships of readability, cloze and multiple choice scores on a reading comprehension test. Journal of Reading Behavior, 10, 417-436.

Erickson, L., & Hansen, L. H. (1974). Performance on cloze passages in and out of context. In P. L. Nacke (Ed.), Interaction: Research and practice for college-age reading. Twenty-third Yearbook of the National Reading Conference. Clemson, South Carolina: NRC.

ETS (1985). TOEFL sample test. Princeton, New Jersey: Educational Testing Service.

Evoldt, C. (1981). A psycholinguistic description of selected deaf children reading in sign language. Reading Research Quarterly, 17, 58-89.

Fillenbaum, S., Jones, L. V., & Rapoport, A. (1963). The predictability of words and their grammatical classes as a function of rate of deletion from a speech transcript. Journal of Verbal Learning and Verbal Behavior, 2, 186-194.

Fischler, I. (1983). Contextual constraint and comprehension of written sentences by deaf college students. American Annals of the Deaf, 128, 418-424.

Flahive, D. E. (1980). Separating the g factor from reading comprehension. In J. W. Oller & K. Perkins (Eds.), Research in language testing. Rowley, Massachusetts: Newbury House.

Fleming, J. T., Ohnmacht, F. W., & Niles, J. A. (1973). Effects of selected strategies and contextual constraint on cloze performance. In P. L. Nacke (Ed.), Diversity in mature reading: Theory and research. Twenty-second Yearbook of the National Reading Conference. Boone, North Carolina: NRC.

Fletcher, J. E. (1960). A study of the relationships between ability to use context as an aid to reading and other verbal abilities. Dissertation Abstracts, 20/07, 2675-2676. (University Microfilm No. 59-5483)

Friedman, M. M. (1964). The use of the cloze procedure for improving the reading comprehension of foreign students at the University of Florida. Dissertation Abstracts, 25/06, 3420-3421. (University Microfilm No. 64-11533)

Fries, C. (1952). The structure of English. New York: Holt, Rinehart & Winston.

Froese, V. (1971). Cloze readability versus the Dale-Chall Formula. (ERIC Document Reproduction Service No. 051 975)

Gallant, R. (1965). Use of cloze tests as a measure of readability in the primary grades. In J. A. Figurel (Ed.), Reading and inquiry. International Reading Association Conference Proceedings, Volume 10. Newark, Delaware: IRA.

Giordano, G. (1985). Learning to read erased text. Academic Therapy, 20, 317-322.

Goodman, K. S. (1971). Psycholinguistic universals in the reading process. In P. Pimsteur & T. Quinn (Eds.), The psychology of second language learning. Cambridge, England: Cambridge University Press.

Gough, P. B. (1983). Context, form and interaction. In K. Raynor (Ed.), Eye movements in reading: Perceptual and language processes. New York: Academic Press.

Greene, F. P. (1965a). A modified cloze procedure for assessing adult reading comprehension. Dissertation Abstracts, 25/10, 5734. (University Microfilm No. 65-5308)

Greene, F. P. (1965b). Modification of the cloze procedure and changes in reading test performances. Journal of Educational Measurement, 2, 213-217.

Guthrie, J. T. (1973). Reading comprehension and syntactic responses in good and poor readers. Journal of Educational Psychology, 65, 294-299.

Hafner, L. E. (1964). Relationships of various measures to the "cloze". In E. L. Thurston & L. E. Hafner (Eds.), New concepts in college-adult reading. Thirteenth Yearbook of the National Reading Conference. Milwaukee, Wisconsin: NRC.

Hanania, E., & Shikhani, M. (1986). Interrelationships among three tests of language proficiency: Standardized ESL, cloze, and writing. TESOL Quarterly, 20, 97-109.

Hanzeli, V. E. (1977). The effectiveness of cloze tests in measuring the competence of students of French in an academic setting. The French Review, 50, 865-873.

Hanzeli, V. E. (1979). Cloze tests in French as a foreign language: Error analysis. In E. J. Briere & F. B. Hinofotis (Eds.), Concepts in language testing: Some recent studies. Washington, DC: TESOL.

Harris, A. J., & Sipay, E. R. (1985). How to increase reading ability: A guide to developmental and remedial methods. White Plains, New York: Longman, Inc.

Harris, D. P. (1985). Some forerunners of cloze procedure. The Modern Language Journal, 69, 367-376.

Hart, B. O., & Rosenstein, J. (1964). Examining the language behavior of deaf children. The Volta Review, 66, 679-682.

Haskell, J. F. (1973a). Refining the cloze procedure for English-as-second-language. The English Review, 25, 77-82.

Haskell, J. F. (1973b). Refining the cloze testing and scoring procedures for use with ESL students. Dissertation Abstracts International, 34/06-A, 3206-3207. (University Microfilm No. 73-31274)

Heilenman, L. K. (1983). The use of a cloze procedure in foreign language placement. The Modern Language Journal, 67, 121-126.

Henk, W. A. (1982). A response to Shanahan, Kamil and Tobin: The case is not yet clozed. Reading Research Quarterly, 17, 591-595.

Hinofotis, F. B. (1980). Cloze as an alternative method of ESL placement and proficiency testing. In J. W. Oller & K. Perkins (Eds.), Research in language testing. Rowley, Massachusetts: Newbury House.

Hinofotis, F. B., & Snow, B. G. (1980). An alternative cloze testing procedure: Multiple-choice format. In J. W. Oller & K. Perkins (Eds.), Research in language testing. Rowley, Massachusetts: Newbury House.

Hisama, K. K. (1977). Patterns of various ESOL proficiency test scores by native language proficiency levels. Occasional papers in linguistics, No. 1. (ERIC Document Reproduction Service No. 144 409)

Hisama, K. K., Lewis, E. L., & Woehlke, P. (1977). A new direction in measuring proficiency in English as a second language. (ERIC Document Reproduction Service No. 150 409)

Horton, J. R. (1973). The construct validity of cloze procedure: An exploratory factor analysis of cloze, paragraph reading, and structure-of-intellect. Dissertation Abstracts International, 34/06-A, 3140. (University Microfilm No. 73-30000)

Horton, J. R. (1974-75). The construct validity of cloze procedure: An exploratory factor analysis of cloze, paragraph reading, and structure-of-intellect tests. Reading Research Quarterly, 10, 248-251.

Irvine, P., Atai, P., & Oller, J. W. (1974). Cloze, dictation, and the Test of English as a Foreign Language. Language Learning, 24, 245-252.

Jenkinson, M. D. (1957). Selected processes and difficulties of reading comprehension. Unpublished doctoral dissertation, University of Chicago.

Jones, M. B., & Pikulski, E. C. (1974). Cloze for the classroom. Journal of Reading, 17, 432-438.

Jongsma, E. A. (1971). The cloze procedure as a teaching technique. Newark, Delaware: International Reading Association.

Jongsma, E. A. (1980). Cloze instruction research: A second look. Newark, Delaware: International Reading Association.

Jonz, J. (1976). Improving on the basic egg: The M-C cloze. Language Learning, 26, 255-265.

Kelly, L., & Ewoldt, C. (1984). Interpreting nonverbatim cloze responses to evaluate program success and diagnose student needs for reading instruction. American Annals of the Deaf, 129, 45-51.

Kennedy, D., & Weener, P. (1973). Visual and auditory training with the cloze procedure to improve reading and listening comprehension. Reading Research Quarterly, 8, 524-541.

Kibby, M. W. (1980). Intersentential processes in reading comprehension. Journal of Reading Behavior, 12, 299-312.

King, C. M., & Quigley, S. P. (1985). Reading and deafness. San Diego, California: College-Hill Press.

Kohler, E. T. (1966). An investigation of cloze scores in terms of selected cognitive variables. Dissertation Abstracts, 27/01-A, 114-115. (University Microfilm No. 66-5454)

LaSasso, C. J. (1978a). A national survey of materials and procedures used to teach reading to hearing impaired children. American Annals of the Deaf, 123, 22-30.

LaSasso, C. J. (1978b). An investigation of the validity and reliability of the cloze procedure as a measure of readability and comprehension for prelingually, profoundly deaf students. Unpublished doctoral dissertation, University of Maryland.

LaSasso, C. J. (1979). The effect of WH question format versus incomplete statement format on deaf students' demonstration of comprehension of text-explicit information. American Annals of the Deaf, 124, 833-837.

LaSasso, C. J. (1980). The validity and reliability of the cloze procedure as a measure of readability for prelingually, profoundly deaf students. American Annals of the Deaf, 125, 559-563.

LaSasso, C. J. (1982). The effectiveness of teacher judgment in determining relative text difficulty for prelingually, profoundly deaf students. Teaching English to Deaf and Second-Language Students. 1, 13-17.

LaSasso, C. J. (1987). The effectiveness of hearing-impaired teachers' judgment of relative text difficulty for deaf students. American Annals of the Deaf, 132, 399-402.

LaSasso, C. J., & Davey, B. (1983a). An examination of hearing impaired readers' test-taking abilities on reinspection tasks. The Volta Review, 85, 279-284.

LaSasso, C. J., & Davey, B. (1983b). An investigation of the criterion-related validity of the SAT-HI Reading Comprehension Subtest for deaf students. In B. B. Bogatz (Ed.), Directions: Preservice education and leadership training. Update on Academic, Professional, Career, and Research Activities, Volume 3, Number 3. Washington, DC: Gallaudet College.

Leys, M., Fielding, L., Herman, P., & Pearson, P. D. (1983). Does cloze measure intersentence comprehension? A modified replication of Shanahan, Kamil, and Tobin. In J. A. Niles & L. A. Harris (Eds.), Searches for meaning in reading/language processing and instruction. Thirty-second Yearbook of the National Reading Conference. Rochester, New York: NRC.

LoMaglio, L. J. (1986). Using cloze type writing exercises to develop functional language competencies in hearing impaired students. Teaching English to Deaf and Second Language Speakers, 1, 16-20.

MacGinitie, W. H. (1961). Contextual constraint in English prose paragraphs. The Journal of Psychology, 51, 121-130.

MacGinitie, W. H. (1964). Ability of deaf children to use different word classes. Journal of Speech and Hearing Research, 7, 33-140.

Manning, W. H. (1986). Cloze-elide: A process oriented model of language proficiency. (ERIC Document Reproduction Service No. 274 669)

Manning, W. H. (1987). Development of cloze-elide tests of English as a second language. TOEFL Research Reports, No. 23. Princeton, New Jersey: Educational Testing Service.

Markham, P. L. (1988). The cloze procedure and intersentential comprehension in college level German. International Review of Applied Linguistics, 26, 44-51.

Marshall, W. A. (1970a). Investigating the effect at fourth grade reading level using the cloze procedure. American Annals of the Deaf, 115, 682-689.

Marshall, W. A. (1970b). Investigating the effect of contextual constraint on deaf and hearing children at the 4th grade reading level, using the cloze procedure. Unpublished doctoral dissertation, University of Illinois, Urbana.

McDaniel, R. R. (1983). A description of the psycholinguistic dimensions of the oral and silent reading processes by third-grade readers of high and average ability and sixth-grade readers of high and low ability. (ERIC Document Reproduction Service No. 228 605)

McLeod, J., & Anderson, J. (1966). Readability assessment and word redundancy of printed English. Psychological Reports, 18, 35-38.

McLeod, J., & Anderson, J. (1970). An approach to assessment of reading ability through information transmission. Journal of Reading Behavior, 2, 116-143.

McNinch, G., Kazelskis, R., & Cox, J. A. (1974). Appropriate cloze deletion schemes for determining suitability of college textbooks. In P. L. Nacke (Ed.), Interaction: Research and practice for college-adult reading. Twenty-third Yearbook of the National Reading Conference. Clemson, South Carolina: NRC.

Meredith, K., & Vaughan, J. (1978). Stability of cloze scores across varying deletion patterns. In P. D. Pearson & J. Hansen (Eds.), Reading: Disciplined inquiry in process and practice. Twenty-seventh Yearbook of the National Reading Conference. Clemson, South Carolina: NRC.

Meyer, R., & Tetrault, E. (1986). Open your CLOZEd minds: using cloze exercises to teach foreign language reading. Foreign Language Annals, 19, 409-415.

Miller, G. R., & Coleman, E. B. (1967). A set of thirty-six prose passages calibrated for complexity. Journal of Verbal Learning and Verbal Behavior, 6, 851-854.

Miller, W. D., & French, S. (1974). Using the cloze procedure to determine the suitability of social science and science textbooks. In P. L. Nacke (Ed.), Interaction: Research and practice for college-adult reading. Twenty-third Yearbook of the National Reading Conference. Clemson, South Carolina: NRC.

Moores, D. F. (1967). Applications of "cloze" procedures to the assessment of psycholinguistic abilities of the deaf. Unpublished doctoral dissertation, University of Illinois, Urbana.

Moores, D. F. (1970). An investigation of the psycholinguistic functioning of deaf adolescents. Exceptional Children, 36, 645-652.

Moores, D. F. (1987a). A new perspective on reading ability of deaf students. Houghton Mifflin / Educators' Forum, Fall, 1,6.

Moores, D. F. (1987b). Educating the deaf: Psychology, principles, and practices. Boston, Massachusetts: Houghton Mifflin Company.

Mullen, K. A. (1979a). An alternative to the cloze test. In C. A. Yorio, K. Perkins, & J. Schachter (Eds.), On TESOL '79: The learner in focus. (ERIC Document Reproduction Service No. 208 683)

Mullen, K. A. (1979b). More on cloze tests of proficiency in English as a second language. In E.J. Briere & F. B. Hinofotis (Eds.), Concepts in language testing: some recent studies. Washington, DC: TESOL.

Nestvold, K. J. (1972). Cloze procedure correlation with perceived readability. Journalism Quarterly, 49, 592-594.

New York City Board of Education (1982). Cloze procedures: A special education teacher's manual. (ERIC Document Reproduction Service No. 255 004)

Nicol, M. A., & Miller, K. M. (1959). Word redundancy in written English. Australian Journal of Psychology, 11, 81-91.

Odom, P., Blanton, R., & Nunnally, J. (1967). Some cloze technique studies of language capability in the deaf. Journal of Speech and Hearing Research, 10, 816-827.

Ohnmacht, F. W., & Fleming, J. T. (1971). Cloze and closure: A second analysis. (ERIC Document Reproduction Service No. 058 015)

Ohnmacht, F. W., Weaver, W. W., & Kohler, E. T. (1970). Cloze and clozure: A factorial study. The Journal of Psychology, 74, 205-217.

Oller, J. W. (1972). Scoring methods and difficulty levels for cloze tests of proficiency in English as a second language. The Modern Language Journal, 61, 151-158.

Oller, J. W. (1973). Cloze tests of second language proficiency and what they measure. Language Learning, 23, 105-118.

Oller, J. W. (1975). Cloze, discourse, and approximations to English. (ERIC Document Reproduction Service No. 107 144)

Oller, J. W. (1979). Varieties of cloze procedure. In J. W. Oller (Ed.), Language Tests at School: A Pragmatic Approach. London: Longman.

Oller, J. W., & Conrad, C. A. (1971). The cloze technique and ESL proficiency. Language Learning, 21, 183-194.

Oller, J. W., & Inal, N. (1971). A cloze test of English prepositions. TESOL Quarterly, 5, 315-326.

Peters, T. E. (1983). The effects of the cloze technique in the measurement of reading comprehension. (ERIC Document Reproduction Service No. 230 903)

Peterson, J., & Carroll, M. (1974). The cloze procedure as an indicator of the instructional level for disabled readers. In P. L. Nacke, (Ed.), Interaction: Research and practice for college-adult reading. Twenty-third Yearbook of the National Reading Conference. Clemson, South Carolina: NRC.

Peterson, J., Paradis, E., & Peters, N. (1973). Revalidation of the cloze procedure as a measure of the instructional level for high school students. In P. L. Nacke (Ed.), Diversity in mature reading: Theory and research. Twenty-second Yearbook of the National Reading Conference, Volume 1. Boone, North Carolina: NRC.

Peterson, J., Peters, N., & Paradis, E. (1972). Validation of the cloze procedure as a measure of readability with high school, trade school, and college populations. In F. P. Greene (Ed.), Investigations relating to mature reading. Twenty-first Yearbook of the National Reading Conference. Milwaukee, Wisconsin: NRC.

Pike, L. W. (1979). An evaluation of alternative item formats for testing English as a foreign language. TOEFL Research Reports, No. 2. Princeton, New Jersey: Educational Testing Service.

Porter, D. (1978). Cloze procedure and equivalence. Language Learning. 28, 333-341.

Porter, D. (1983). The effect of quantity of context on the ability to make linguistic predictions: A flaw in a measure of general proficiency. In A. Hughes & D. Porter (Eds.), Developments in language testing. London, England: Academic Press.

Potter, T.C. (1968). A taxonomy of cloze research. Part I: Readability and reading comprehension. (ERIC Document Reproduction No. 035 514)

Propst, I. K., & Baldauf, R. B. (1979). Use matching cloze tests for elementary ESL students. The Reading Teacher, 32, 683-690.

Quigley, S. P., & Paul, P. V. (1984). Language and deafness. San Diego, California: College-Hill Press.

Raatz, U. (1985). Tests of reduced redundancy - The C-test, a practical example. (ERIC Document Reproduction Service No. 266 684)

Raatz, U., & Klein-Braley, C. (1981). The C-test - a modification of the cloze procedure. In T. Culhane (Ed.), Proceedings of the international language testing symposium of the Interuniversitare Sprachtestgruppe. (ERIC Document Reproduction Service No. 217 724)

Ragosta, M., & Nelson, C. (1986). TOEFL and hearing impaired students: A feasibility study. (ERIC Document Reproduction Service No. 269 422)

Ramanauskas, S. (1972). The responsiveness of cloze readability measures to linguistic variables operating over segments of text longer than a sentence. Reading Research Quarterly, 8, 72-91.

Rankin, E. F. (1958). An evaluation of the cloze procedure as a technique for measuring reading comprehension. Dissertation Abstracts, 19/04, 733-734. (University Microfilm No.1 58-3722)

Rankin, E. F. (1959a). The cloze procedure - its validity and utility. In O. S. Causey & W. Eller (Eds.), Starting and improving college reading programs. Eighth Yearbook of the National Reading Conference. Fort Worth, Texas: Texas Christian University.

Rankin, E. F. (1959b). Uses of the cloze procedure in the reading clinic. In J. A. Figurel (Ed.), Reading in a changing society. International Reading Association Conference Proceedings, Volume 4. New York: Scholastic Magazines.

Rankin, E. F. (1965). The cloze procedure - a survey of research. In E. L. Thurston & L. E. Hafner (Eds.), The philosophical and sociological bases of reading. Fourteenth Yearbook of the National Reading Conference. Milwaukee, Wisconsin: NRC.

Rankin, E. F. (1974). The cloze procedure revisited. In P. L. Nacke & E. Neill (Eds.), Interaction: Research and practice for college-age reading. Twenty-third Yearbook of the National Reading Conference. Clemson, South Carolina: NRC.

Rankin, E. F. (1977). Sequence strategies for teaching reading comprehension with the cloze procedure. In P. D. Pearson (Ed.), Reading: theory, research, and practice. Twenty-sixth Yearbook of the National Reading Conference. Clemson, South Carolina: NRC.

Rankin, E. F. (1978). Characteristics of the cloze procedure as a research tool in the study of language. In P. D. Pearson & J. Hansen (Eds.), Reading: Disciplined inquiry in process and practice. Twenty-seventh Yearbook of the National Reading Conference. Clemson, South Carolina: NRC.

Rankin, E. F., & Culhane, J. W. (1969). Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 13, 193-198.

Rankin, E. F., & Helm, P. (1986). The validity of cloze tests in relation to a psycholinguistic conceptualization of reading comprehension. (ERIC Document Reproduction Service No. 272 841)

Rankin, E. F., & Thomas S. (1980). Contextual constraints and the construct validity of the cloze procedure. In M. L. Kamil & A. J. Moe (Eds.), Perspectives on reading research and instruction. Twenty-ninth Yearbook of the National Reading Conference. Washington, D.C.: NRC.

Ransom, P. E. (1966). A study to determine reading levels of elementary school children by cloze testing. Dissertation Abstracts, 26/07, 3705-3706. (University Microfilm No. 65-13038)

Ransom, P. E. (1968). Determining reading levels of elementary school children by cloze testing. In J. A. Figurel (Ed.), Forging Ahead in Reading. Proceedings of the Twelfth Annual Convention, International Reading Association, Part 1. Newark, Delavare: IRA.

Riley, J. D. (1986). Progressive cloze as a remedial technique. The Reading Teacher, 39, 576-581.

Robinson, R. D. (1972). An investigation into the use of the cloze procedure with a group of functionally illiterate adults. Dissertation Abstracts International, 32/07-A, 3572. (University Microfilm No. 72-2535)

Ruddell, R. B. (1964a). An investigation of the effect of the similarity of oral and written patterns of language structure on reading comprehension. Dissertation Abstracts, 24/12, part 1, 5207-5208. (University Microfilm No. 64-3826)

Ruddell, R. B. (1964b). A study of the cloze comprehension technique in relation to structurally controlled reading material. In J. A. Figurel (Ed.), Improvement of reading through classroom practice. International Reading Association Conference Proceedings, Volume 9. Newark, Delavare: IRA.

Rye, J. (1984). Cloze and intersentential constraint. Journal of Research in Reading, 7, 113-122.

Salzinger, K., Portnoy, S., & Feldman, R. S. (1962). The effect of order of approximation to the statistical structure of English on the emission of verbal responses. Journal of Experimental Psychology, 64, 52-57.

Sampson, M. R., Valmont, W. J., & Van Allen, R. (1982). The effects of instructional cloze on the comprehension, vocabulary, and divergent production of third-grade students. Reading Research Quarterly, 17, 389-399.

Schneyer, J. W. (1965). Use of the cloze procedure for improving reading comprehension. The Reading Teacher, 19, 174-179.

Schoelles, I. S. (1971). Cloze as a predictor of reading group placement: Research report. (ERIC Document Reproduction Service No. 053 868)

Scholz, G. E., & Scholz, C. M. (1981). Multiple-choice cloze tests of EST discourse: An exploration. (ERIC Document Reproduction Service No. 208 656)

Shanahan, T., & Kamil, M. L. (1982). The sensitivity of cloze to passage organization. In J. A. Niles & L. A. Harris (Eds.), New inquiries in reading research and instruction. Thirty-first Yearbook of the National Reading Conference. Rochester, New York: NRC.

Shanahan, T., & Kamil, M. L. (1983). A further comparison of sensitivity of cloze and recall to passage organization. In J. A. Niles & L. A. Harris (Eds.), Searches for meaning in reading/language processing and instruction. Thirty-second Yearbook of the National Reading Conference. Rochester, New York: NRC.

Shanahan, T., Kamil, M. L., & Tobin, A. W. (1982). Cloze as a measure of intersentential comprehension. Reading Research Quarterly, 17, 229-255.

Simons, H. D. (1970). The relationship between aspects of linguistic performance and reading comprehension. Unpublished doctoral dissertation, Harvard Graduate School of Education.

Simons, H. D. (1971). Reading comprehension: The need for a new perspective. Reading Research Quarterly, 6, 338-363.

Smith-Burke, M., Gingrich, P. S., & Eagleeye, D. (1978). Differential effects of prior context, style and deletion pattern on cloze comprehension. In P. D. Pearson & J. Hansen (Eds.), Reading: Disciplined inquiry in process and practice. Twenty-seventh Yearbook of the National Reading Conference. Clemson, South Carolina: NRC.

Spolsky, B. (1969). Reduced redundancy as a language testing tool. (ERIC Document Reproduction Service No. 031 702)

Stubbs, J. B., & Tucker, G. R. (1974). The cloze test as a measure of English proficiency. The Modern Language Journal, 58, 239-241.

Suhorsky, J. (1975). An investigation of the relationship between undeleted text preceding a cloze test and cloze test results. Dissertation Abstracts International, 36/06-A, 3393. (University Microfilm No. 75-27797)

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. Journalism Quarterly, 30, 415-433.

Taylor, W. L. (1956). Recent developments in the use of "cloze procedure". Journalism Quarterly, 33, 42-48.

Taylor, W. L. (1957). "Cloze" readability scores as indices of individual differences in comprehension and aptitude. Journal of Applied Psychology, 41, 19-26.

Tuinman, J. J. (1972). Understanding comprehension: Aspects of its measurement. In L. A. Harris & C. B. Smith (Eds.), Individualizing reading instruction. New York: Holt, Rinehart and Winston.

Valmont, W. J. (1983). Instructional cloze procedures: Rationale, framework, and examples. Reading Horizons, 23, 156-162.

Wainman, H. (1979). Cloze testing of second language learners. English Language Teaching, 33, 126-132.

Wardell, D. (1985). Cloze procedure, An historical review. (ERIC Document Reproduction Service No. 264 306)

Weaver, W. W., & Bickley, A. C. (1967). Structural-lexical predictability of materials which predictor has previously produced or read. In E. Mallinoff (Ed.), Proceedings of the 75th annual convention of the American Psychological Association, Volume 2. Washington, DC: American Psychological Association.

Weaver, W. W., & Kingston, A. J. (1963). A factor analysis of the cloze procedure and other measures of reading and language ability. Journal of Communication, 13, 252-261.

Webster, A., Wood, D. J., & Griffiths, A. J. (1981). Reading retardation or linguistic deficit? I: Interpreting reading test performance of hearing-impaired adolescents. Journal of Research in Reading, 4, 136-147.

Willows, D. M., & Ryan, E. B. (1981). Differential utilization of syntactic and semantic information by skilled and less skilled readers in the intermediate grades. Journal of Educational Psychology, 73, 607-615.

Willows, D. M., & Ryan, E. B. (1986). The development of grammatical sensitivity and its relationship to early reading achievement. Reading Research Quarterly, 21, 253-266.

Wilson, K., Karchmer, M., & Jensema, C. (1980). Literal vs. inferential item analysis of reading achievement in hearing impaired students. In H. N. Reynolds & C. M. Williams (Eds.), Proceedings of the Gallaudet conference on reading in relation to deafness. Washington, DC: Gallaudet College Division of Research.

Wolk, S., & Schildroth, A. (1984). Consistency of an associational strategy used by hearing-impaired students. Journal of Research in Reading, 7, 135-142.

Wood, D. J., Griffiths, A. J., & Webster A. (1981). Reading retardation or linguistic deficit? II: Test-answering strategies in hearing and hearing-impaired school children. Journal of Research in Reading, 4, 148-156.

APPENDIX

TABLE 1. Results of ranking studies

Study	Subjects	Ranking Criteria	Agreement with cloze
Taylor '53	college	Dale-Chall, Flesch formulae	same ranking
Bormuth '63a	4th, 5th, 6th grades	Dale-Chall	same ranking
		multiple-choice performance	.92
Bormuth '68	4th, 5th, 6th grades	multiple-choice performance	.90-.96
Froese '71	6th grade	Dale-Chall	-.29
Anderson '72	*primary	teacher judgment	same ranking
Nestvold '72	college	SMOG formula	.47
Oller '72	*college	Dale-Chall, Flesch formulae	same ranking[1]
Robinson '72	adult ("functionally illiterate")	readability formulae[2]	n.s.
Haskell '73a,b	*adult	Thorndike list	same ranking
Entin & Klare '78	college	Flesch formula	.68
		multiple-choice performance	n.s.
Alderson '80	secondary, *college	teacher judgment, Dale-Chall and other formulae	same ranking

*ESL students
[1]This finding is for exact word scoring. Under acceptable alternate scoring, cloze performance did not discriminate between two of the three passages.
[2]Specific formulae are not reported in Dissertation Abstracts.

TABLE 2. Correlations of cloze performance by native speakers with same-passage multiple-choice measures

Study	Subjects	Results[1]
Taylor '56	adults	.76
Jenkinson[2] '57	secondary	.82
Rankin '59a	college	.78
Bormuth '63a	4th, 5th, 6th grades	.73-.84 .93 (total)
Bormuth '67	4th, 5th grades	.95
Rankin & Culhane '69	5th grade	.68
Entin & Klare '78	college	.57

[1]Results are given for any-word deletion and "pre-cloze" administration (i.e., cloze test administered before reading the passage) for all studies except Rankin.
[2]Statistics reported in Rankin 1959a

TABLE 3. Correlations of cloze performance by native speakers with standardized reading measures

Study	Subjects	Reading Test	Results[1]
Jenkinson[2] '57	secondary	CRT[3]	.73
Rankin '59a	college	DRT[4]	.29-.60
Fletcher '60	college	CRT[3]	.55
Weaver & Kingston '63	college	Davis[5]	.25-.51
Ruddell '64b	4th grade	SAT[6]	.61-.72
Gallant '65	1st, 2nd, 3rd grades	MAT[7]	.65-.81
Greene '65b	college	DRT[8]	.51
Schneyer '65	6th grade	GRS[9]	.60
Froese '71	6th grade	CTBS[10]	.52-.85
Schoelles '71	1st-5th grades	SAT[11]	.64-.82

[1]Results are given for exact word scoring and any-word deletion, where more than one method was employed.
[2]Statistics reported in Rankin 1959a
[3]Cooperative Reading Test: Level of Comprehension
[4]Dignostic Reading Test-Survey Section: Story Comprehension (.29) and Paragraph Comprehension (.60)
[5]Davis Reading Test: total score
[6]Stanford Achievement Test: Paragraph Meaning Section
[7]Metropolitan Achievement Test: Reading Test
[8]Diagnostic Reading Test-Survey Section: total comprehension
[9]Gates Reading Survey
[10]Canadian Test of Basic Skills: Comprehension Subtest
[11]Stanford Achievement Test: Word Meaning Section and Paragraph Meaning Section combined

TABLE 4. Correlations of cloze performance by second language learners with reading comprehension measures

Study	Subjects[1]	Reading Test[2]	Results EX[3]	Results AC[4]
Friedman[5] '64	college	MAT	.71-.87	
Oller & Conrad '71	college	UCLA	.80	
Oller '72	college	UCLA	.65-.77	.71-.82
Irvine, et al. '74	adults	TOEFL	.67	.65
Stubbs & Tucker '74	adults	EEE	.67	.70
Alderson '79b	adults	ELBA	.47-.54	.58-.63
Pike '79	adults	TOEFL	.65-.79	
Flahive '80	college	M-H		.65
		P-Y		.62
Hinofotis '80	college	TOEFL	.68	.77
		SIU	.80	.80
Bachman '85	college	TOEFL		.80

[1]All subjects were ESL students.
[2]Reading Comprehension subtests from ESL placement exams used at UCLA, Southern Illinois University (SIU), American University of Beirut (EEE), in Great Britain (ELBA), and from the Test of English as a Foreign Language (TOEFL). Also, standardized tests of reading comprehension for native speakers: Metropolitan Achievement Test, Reading Section (MAT), McGraw-Hill Basic Skills System Reading Test (M-H), and the Perkins-Yorio Test (P-Y).
[3]Exact word scoring
[4]Acceptable alternate scoring
[5]Statistics reported in Rankin 1965.

TABLE 5. Correlations of cloze performance by second language learners with language proficiency measures

Study	Subjects[1]	Test[2]	Results	
			EX[3]	AC[4]
Carroll *et al.* '59	secondary	CEEB	.62	
Oller & Conrad '71	college	UCLA	.88	
Oller '72	college	UCLA	.73-.87	.80-.89
Irvine *et al.* '74	adults	TOEFL	.78	.79
Stubbs & Tucker '74	adults	EEE	.71	.76
Alderson '79b	adults	ELBA	.61-.67	.71-.80
Briere & Hinofotis '79	college	UCLA	.83	
		USC	.78	
		SIU	.80	
Mullen '79b	college	TOEFL	.55-.66	.62-.70
Flahive '80	college	TOEFL		.75
Hinofotis '80	college	TOEFL	.71	.79
		SIU	.80	.84
Caulfield & Smith '81	secondary	MLAS		.90
Heilenman '83	college	MLAF		.94
Bachman '85	college	TOEFL		.83
Hanania & Shikhani '86	college	AUB	.79	

[1]All subjects were ESL students, with three exceptions: subjects in the Carroll *et al.* and Heilenman studies were English speaking students learning French, and in Caulfield & Smith's study they were English speaking sophomores learning Spanish.
[2]ESL placement tests from UCLA, USC, Southern Illinois University (SIU), American University of Beirut (EEE and AUB), and from Great Britain (ELBA). Also, foreign language proficiency measures: College Entrance Examination Board language test-French (CEEB), Test of English as a Foreign Language (TOEFL) and Modern Language Association Cooperative Test-Spanish (MLAS) and -French (MLAF).
[3]Exact word scoring
[4]Acceptable alternate scoring

TABLE 6. Results of criterion cloze score studies

Study	Subjects	Results	
		Cloze	M-C Test
Regression:			
Bormuth '67	4th, 5th grades	42%-53%[1]	75%-90%
Alexander '68	'intermediate' grades	47%-61%	75%-90%
Rankin & Culhane '69	5th grade	41%-61%	75%-90%
Jones & Pikulski '74	6th grade	36%-43%	75%-90%
Entin & Klare '78	college	31%-36%[2]	75%-90%
Matching:			
Bormuth '68	4th, 5th, 6th grades	44%-57%	75%-90%
Almeida '75	6th grade	40%	70%[3]
		40%	70%[4]
		40%	77%[5]
Grade Level Performance:			
Peterson *et al.* '72, '73	high school-	42%	
	adult vocational-	43%	
	college level-	44%	
Peterson & Carroll '74	6th, 7th, 8th, 9th grades (disabled readers)	41%	
Maximum Gain:			
Bormuth '69a	3rd, 5th, 7th, 11th grades, jr. college, grad students	37%-56%	
Method Not Reported:			
Anderson '71	*8th grade	44%-53%	75%-90%

*ESL students
[1]corrected for guessing on multiple-choice test
[2]lookback on m-c test yielded artificially lov cloze scores
[3]m-c score of S's scoring 40% on 4th grade cloze test
[4]m-c score of S's scoring 40% on 5th/6th grade level cloze
[5]m-c score of S's scoring 40% on 7th/8th grade level cloze